Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

Short-Term Retention as a Function of the Average Number of Items Presented: K. E. LLOYD, L. S. REID, AND J. B. FEALLOCK 201

Response Strength as a Function of Changed Intertrial Interval: C. B. ERNHART 208

Poststimulus Cuing in Immediate Memory: N. S. ANDERSON 216

The Immediate Retention of Unrelated Words: B. B. MURDOCK, JR. 222

Anisotropy and Interaction of Fields of Spatial Induction: R. M. MICHAELS 235

Reaction Time as a Function of Stimulus Information and Age: G. J. SUCI, M. D. DAVIDOFF, AND W. W. SURWILLO 242

Retroactive Inhibition of Connected Discourse as a Function of Similarity of Topic: N. J. SLAMECKA 245

Shock Strength, Shock Reduction, and Running Speed: J. P. SEWARD, R. A. SHEA, A. A. UYEDA, AND D. C. RASKIN 250

Influence of Extratest Illumination on the Critical Flicker Frequency of the Human Fovea: H. RIPPS AND I. T. KAPLAN 255

Supplementary Report: Effectiveness of Food and Nonfood Signs in Reversal Learning in Monkeys: J. M. WARREN 263

American Psychological Association

Vol. 60, No. 4 October 1960

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association.*** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1960 the publication cost is $30 per printed page.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Promotion Manager
BEATRICE B. HAMLIN, Editorial Assistant

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1960 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 60, NO. 4 OCTOBER 1960

SHORT-TERM RETENTION AS A FUNCTION OF THE AVERAGE NUMBER OF ITEMS PRESENTED [1]

KENNETH E. LLOYD,[2] LYNE STARLING REID, AND JOHN B. FEALLOCK [3]

University of Virginia

In comparison to long-term retention, research on short-term retention has lagged far behind with experiments limited for the most part to the classic memory span. Such studies involve presenting *S* a list of items, and requiring him to recall the items in some specified order. The experiments to be reported attempt an analysis of short-term retention in a different kind of experimental situation, one in which the presentation of items and their recall are not discretely separated. Instead, both occur in a variety of serial orders over an extended period of time.

In this task each item to be remembered is followed at some point in the sequence by a cue which specifies that *S* must recall that item. If the items to be remembered are viewed as instances of a class then the class name can be used as a recall point (e.g., if the word, "pine," is a stored item then the word, "TREE" serves as the recall point). A sequence is initiated by presenting *S* with an item to be remembered. The recall point for this item may follow immediately in the sequence, or may be separated from it by several other items, by other recall points, or by both. At the time a given item occurs *S* may be remembering several other items; similarly, at any recall point *S* may be remembering various numbers of prior items. Throughout the task *S* is being presented with new information which he must store, and is being requested to recall information which he has been storing.

Table 1 contains three short sequences as illustration. The lower case letters represent items to be remembered, or instances of a class, such as "pine," or "tin," as shown in the third sequence. The upper case letters indicate their respective recall points such as "TREE" or "METAL." Inspection of these sequences suggests the variety of serial orders that is possible.

In the top sequence in Table 1 *S* is presented with *a*, *b*, and *c* which he successively stores. At *B* he is requested to recall *b*. The bottom line in Table 1 is a repetition of the initial portion of the top sequence with words substituted for the letters. In the example just given *S* is storing "pine," "tin," and "polo," and when he is presented the class word, "METAL," he must recall

[1] The authors wish to thank William F. Hawkins who performed much of the work that made these studies possible. The studies were supported by Contract No. DA-49-007-MD-537 between the University of Virginia and the Office of the Surgeon General, Department of the Army for which Richard H. Henneman is principal investigator.

[2] Now at Washington State University.

[3] Now at the Ohio State University.

TABLE 1

SAMPLE SUB-SEQUENCES WITH DIFFERENT AVERAGE STORAGE LOADS
(Lower case letters represent items to be stored and upper case letters indicate their respective recall points.)

	Serial Position	Average Storage Load
	a b c B b A B d C e f E D d g h D e F c G H E C	
Storage Load	3 3 2 2 3 2 4 4 4 3 2 1	2.8
	a b c d B e f g h D b G d E e A H F C h B E D H	
Storage Load	4 7 7 7 7 6 5 4 4 3 2 1	4.8
	pine tin polo METAL copper TREE METAL hat SPORT dog . . .	
Storage Load	3 3 2 2	

"tin." The variables involved in the retention of *a*, *b*, and *c* can be analyzed in several ways. When *b* occurs *S* is already storing *a*; with reference to *b*, the *a* may be termed a prior item. When *c* occurs *S* is storing *a* and *b*; again, with reference to *b*, the *c* may be labeled an interpolated item, i.e., *c* intervenes between *b* and *B*. Another kind of interpolation occurs with the *b A B* in the top sequence. Here, a recall point, *A*, is interpolated between *b* and *B*. These arrangements occur in a variety of patterns throughout any sequence, and by analogy to long-term retention paradigms, can be thought of as having proactive and retroactive effects.

Within the sequences described, however, prior and interpolated items and recall points cannot be manipulated independently of each other. For example, a failure to recall *b* in the *b A B* situation cannot be attributed solely to the interpolated *A* since there must be a prior item in the sequence that *S* is storing until *A* occurs. In other words, the *b A B* can occur in various contexts within any sequence ranging from *a b A B* to something like *d a c C b A B*.

This nonindependence of conditions makes highly desirable the use of variables which summarize these dependencies within the task. The most obvious possibility is the average number of items *S* is required to remember over a sequence. The value of this variable, average storage load, can be obtained by adding the loads at each recall point and dividing the sum by the number of recall points. In Table 1 the numbers beneath each recall point designate the storage load at that recall point. For example, at *B* the *S* is storing *a*, *b*, *c*, and the load at *B* is 3; at *A* the load is again 3 as *S* is storing *a*, *c*, and the second *b*. The average storage loads for the sequences in Table 1 are shown in the right-hand column.

The adequacy of a variable of this sort depends upon the extent to which it can be shown to be related to performance under a variety of conditions. One test is the extent to which recall errors increase directly with increases in average storage load. Exp. I was designed for this purpose.

Another test is the extent to which performance remains unaltered when storage load is held constant, despite changes in other conditions in the task. One such change might involve the number of different items *S* encounters throughout the task. Since the items stored have long been a part of his response repertoire, performance may well be independent of the fact that under one condition the same items are repeated, while under another condition different items always occur. The size of a population of items appearing in a sequence can be varied in at least two ways: (*a*) the number of different

item classes in the sequence can vary while the number of items within a class is held constant; (*b*) the number of items within a class can vary while the number of classes is constant. Exp. II and III attempt to determine whether, when varied in these two ways, the population of items in a sequence influences performance.

In the experiments to be reported, not only are the sequences used a sample from a population of sequences, but the task itself is only one of many in which average storage load could be an important variable. In the present task *S* is never required, at any one time, to store and recall more than one item from each class. But the sequences can be altered so that *S* is required to store varying numbers of items from each class; and so have to recall one or more of the stored items at the various recall points. Other task variations are possible. It is obvious that further tests of the effectiveness of average storage load in predicting performance need to be made under a variety of conditions.

Experiment I

Method

In Exp. I the average number of items that *Ss* were required to remember was systematically varied. The *Ss* stored and recalled words presented to them aurally by a magnetic tape recorder. They listened to sequences composed of 12 sub-sequences constructed to have any one of four values of average storage load (1.8, 2.8, 3.8, and 4.8). Within a sequence each value of average load appeared three times in a partially counterbalanced order. Since a sub-sequence contained 12 items and 12 recall points, 144 items were presented for recall in a sequence. Table 1 contains two of these sub-sequences.

To increase the range of sub-sequences sampled, three different sub-sequences were constructed for each value of average storage load. They differed from each other in terms of which items were requested at specific recall points, e.g., if *S* is storing *a*, *b*, and *c* then the first recall point following *c* could be *A*, *B*, or *C*, and the storage load for that recall point would not change. The serial position of recall points in the three sub-sequences of a given average storage load remained the same, but the particular class name that occurred at a given recall point was changed to produce these variations.

The items were instances of a class of English words, and the class names were used as recall points. Eight different word classes were used: automobile, tree, city, metal, sport, clothing, color, and animal. The four items in each class were judged to be familiar members of that class. The automobiles, for example, were Buick, Dodge, Ford, and Nash; the metals were brass, copper, iron, and tin. The words were probably not equally familiar to the *Ss*, but over the sequences each word occurred in many different load conditions. Six sequences were recorded on the magnetic tape. The serial order of the 12 sub-sequences differed in each sequence. A presentation rate of one item or recall point every 3 sec. was used. Recall points were preceded by a short, 500-cps tone to aid *S* in detecting them.

The 10 *Ss*, University of Virginia undergraduates, were briefly shown a list of the 32 words grouped into the eight classes. They were instructed to remember each item as they heard it on the tape, and to recall the items when the appropriate class name was heard. They were encouraged to guess if uncertain about a recall. Each of the 10 *Ss* served 1 hr. a day on four consecutive days. On Day 1 *Ss* responded to two or three sequences depending upon the time needed for instructions. On the remaining three days *Ss* responded to three sequences each day. All responses were recorded.

Results

Except for Day 1 when some *Ss* responded to only six, each *S* responded daily to nine sub-sequences at each value of average storage load. An average of his errors over these repeated measures was obtained. Table 2 presents the mean errors for all *Ss* over the four days of practice. It will be noted that recall errors increase in a systematic manner with increases in average load.

TABLE 2

MEAN ERRORS AS A FUNCTION OF AVERAGE STORAGE LOAD

Day	Average Storage Load							
	1.8		2.8		3.8		4.8	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	1.20	.60	2.68	1.06	4.64	1.05	6.01	1.12
2	.77	.57	2.05	1.38	3.55	.88	4.92	1.40
3	.59	.40	1.51	1.05	3.09	1.23	4.93	1.57
4	.35	.53	1.18	1.25	2.74	1.61	3.99	1.47

The recall scores on Day 4 are considered to be the best estimate of stable performance. Table 3 contains a trend analysis (Grant, 1956) of the average errors for each *S* on this day. For this analysis a square-root transformation of the data was made since the errors for loads 1.8 and 2.8 are skewed. From Table 3 it can be seen that not only is the over-all trend *F* ratio highly significant, but that the linear component accounts for almost 100% of the variance attributed to this trend.

A second analysis of variance was made for all days of practice. Average storage load was again highly significant, as was the effect of practice. The Load × Practice interaction was not significant.

TABLE 3

TREND ANALYSIS OF TRANSFORMED SCORES OF AVERAGE STORAGE LOAD

Source	*df*	*MS*	*F*
Over-all trend	3	2.84	54.61*
Linear	1	8.44	129.84*
Quadratic	1	0.00	—
Cubic	1	0.06	1.93
Between-individual means	9	0.49	9.42*
Between-individual trends	27	0.05	
Linear	9	0.07	
Quadratic	9	0.06	
Cubic	9	0.03	
Total	39		

* $P < .01$.

EXPERIMENT II

Method

This study and Exp. III further test the adequacy of average storage load in predicting performance on a sequential task. In these studies average storage load was held constant while the size of the population of items used to construct the sequences was varied. In Exp. II this was accomplished by varying the number of item classes used in a sequence. Four groups of *Ss* responded to sequences which contained items from either 6, 8, 10, or 12 classes. The number of items in a class was always four. It was conjectured that if the items are relatively familiar, the size of the population from which they are drawn does not influence the number of recall errors when average storage load is held constant. The 12 class names were automobile, tree, city, color, sport, college, metal, bird, relative, water, animal, and clothing.

All word classes were used in the sequences for the group responding to 12 classes. In order to control for possible variations due to the particular words, the groups responding to 6, 8, and 10 classes were each divided into two subgroups which differed only in the words used. Inspection of the recall errors indicated that none of these subdivisions produced any performance differences.

The five sub-sequences used in Exp. II were similar to those in Exp. I, but contained only 10 items and 10 recall points. These five sub-sequences varied from 3.3 to 3.6 in average storage load. Each of the five occurred three times per sequence for a total of 150 items for recall.

Four different sequences were recorded on magnetic tape for each group in order to balance the serial order of sub-sequences and the particular words used at the recall points. A presentation rate of one item or one recall point every 1.5 sec. was used.

This is shorter than the 3-sec. rate of Exp. I and III.

Twenty-four University of Virginia undergraduates, six in each group, responded to three different sequences each day for four days. The instructions were the same as those in Exp. I.

Results

On each day *S* responded to 45 sub-sequences. The errors on these repeated measures were averaged for each *S*, and the mean errors of the four groups are presented in Table 4. An analysis of variance of errors on Day 4 results in an *F* ratio of less than one. Over the four days of practice only the effect of practice was significant. The number of item classes in a sequence did not alter recall at any stage of practice on the sequential task.

The average storage loads for the five different sub-sequences in each sequence were 3.3, 3.3, 3.3, 3.5, and 3.6; the mean errors for all 24 *S*s on Day 4 for these five sub-sequences were 3.49, 3.61, 3.46, 3.85, and 4.50, respectively. Even though the values of storage load were closely spaced, remarkable agreement was maintained between recall error and average load.

Experiment III

Method

Experiment III is another test of the effect of varying the size of the population from which the items appearing in a sequence are drawn. In contrast to Exp. II the number of different items in each class was varied while holding constant the number of classes. Varying the number of items in a class presents a problem, however, for the possibility exists that differences may occur merely as a result of different opportunities for guessing. If the number of items is small the likelihood of correct recalls by chance is greatly increased. For this reason it was decided to use classes containing 8 and 16 items. These values should result in a chance error difference of less than one over 12 recall points.

The effects of number of items and average storage load were tested in a factorial design. The number of items was varied as a within-*S* condition. This design was chosen so that all *S*s encountered the words the same number of times. All sequences contained items from six classes, but one-half the sequences contained 8 items from each class while the others contained 16 items.

Average storage load was varied as a between-*S* variable. The *S*s responded to sequences with an average storage load of 2.8, and 10 other *S*s to sequences with an average load of 3.8.

The sub-sequences from Exp. I were used. Four variations of each sub-sequence were repeated twice to form a total sequence of 96 items to be recalled. If the sequence contained 8 items per class each item was repeated twice; in 16-item sequences each item appeared only once. Three sequences were recorded for each of the four conditions to balance the words, and the serial order of sub-sequences. The 3-sec. presentation rate of Exp. I was used.

Instructions to the 20 University of Virginia undergraduates were the same as in previous experiments. The *S*s responded to one or two sequences on Day 1 depending upon how much time was devoted to instruc-

TABLE 4

Mean Errors as a Function of Number of Item Classes: Exp. II

Day	Number of Item Classes							
	6		8		10		12	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	5.69	.55	5.04	.42	5.14	.63	5.15	.73
2	4.83	.50	4.24	.56	4.52	.50	4.32	.49
3	4.50	.64	3.86	.70	4.09	.43	3.97	.51
4	4.01	.85	3.55	.75	3.82	.33	3.62	.35

TABLE 5

MEAN ERRORS AS A FUNCTION OF THE NUMBER OF ITEMS AND AVERAGE STORAGE LOAD: EXP. III

Day	Low Storage Load				High Storage Load			
	8 Items		16 Items		8 Items		16 Items	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	2.61	1.19	3.37	1.31	5.71	1.52	5.94	1.83
2	1.95	1.32	2.34	1.29	5.19	1.39	5.45	1.54
3	1.81	1.10	2.09	1.29	4.49	1.72	5.07	1.46
4	1.58	1.19	1.61	1.15	4.10	1.71	4.16	1.87

tions. On each of the last three days of practice they responded to four different sequences. In two of these sequences the classes contained 8 items and in the other two 16 items.

Results

All *S*s responded on each day to sub-sequences under both the 8- and 16-item conditions. The errors on these repeated measures were averaged for each *S*, and the means and *SD*s for all conditions over the four days are presented in Table 5. The means for the high and low storage loads are for different *S*s. A summary of the analysis of variance for these means is presented in Table 6. Load and Practice are significant at the 0.1% level, Items at the 0.5% level, but none of the interactions is significant. It will be noted in Table 5 that the differences between the 8- and 16-item conditions under the two storage loads have disappeared on Day 4. When the data for this day were subjected to an analysis of variance only average storage load emerged as a significant effect. However, the absence of a significant Items × Practice interaction over the four days makes it difficult to interpret the Day 4 analysis. It is quite possible that with several additional days of practice such an interaction would have been demonstrated, but in light of the results it is necessary to leave open the possibility that number of items contributes a slight but significant effect, even when performance has stabilized.

The 16-item condition required *S* to store a different item every time throughout the sequence. This would seem to be an adequate upper limit for testing this variable. If fewer than eight alternatives had been selected, then performance would have been enhanced by an increase in the number of correct guesses. A comparison of the errors in the second and third columns of Table 2

TABLE 6

ANALYSIS OF VARIANCE OF NUMBER OF ITEMS AND AVERAGE STORAGE LOAD OVER FOUR DAYS OF PRACTICE: EXP. III

Source	*df*	*MS*	*F*
Between-*S*s	19		
Load	1	323.91	20.58***
Error	18	15.74	
Within-*S*s	140		
Items	1	4.27	10.17**
Practice	3	16.95	40.36***
I × P	3	0.38	
I × L	1	0.07	
P × L	3	0.68	
I × P × L	3	0.30	
Error	126	0.42	
Total	159		

** $P < .005$.
*** $P < .001$.

with those of Table 5 indicates fewer errors with the four alternative subsequences in Exp. I. However, when the scores in Exp. I are corrected for chance, the differences disappear.

Discussion

For the experimental arrangement investigated, these three experiments seem to establish the fact that the average number of items *Ss* are required to remember is related in a reliable manner to their recall errors. This proves to be the case despite wide variation in the population of items being stored. For average storage load is significantly related to performance whether items are varied in terms of the number of classes or the number of items within a class. Pollack, Johnson, and Knaff (1959) report a similar finding in the running memory span when they varied the size of the population from which the items to be recalled were drawn.

The direction that further work must take is obviously one of assessing the generality of the notion of average storage load. This can be done by studying this variable in sequential tasks that differ from the present one. One such variation is to depart from the one-to-one relationship between stored items and recall points. As mentioned in the introduction, *S* can be required to recall one or more items at each recall point. In such a task further variations can be introduced by requiring *Ss* to recall related items from fixed classes, as in the present experiments, or unrelated items that are arbitrarily shifted from class to class. In the latter instance the classes can be designated by some symbol such as a letter, and the items shifted from one letter class to another as they are presented for storage at different points in the sequence. The importance of average storage load for short-term retention will be increased greatly with the demonstration that it determines performance under these circumstances.

Summary

Short-term retention was studied in a situation requiring *S* to remember familiar English words until their recall was requested. During the time *S* was required to remember these items he was presented other words or asked to recall previously presented items. Three experiments are reported in which the first demonstrated that recall errors increased systematically as the average number of words to be remembered over blocks of recall points, termed average storage load, was increased. The other two experiments further demonstrated that performance was related to average storage load, but was relatively independent of the size of population from which the English that made up the tasks was drawn.

References

Grant, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

Pollack, I., Johnson, L. B., & Knaff, P. R. Running memory span. *J. exp. Psychol.*, 1959, **57**, 137–146.

(Received October 7, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 4, 208–215

RESPONSE STRENGTH AS A FUNCTION OF CHANGED INTERTRIAL INTERVAL [1]

CLAIRE B. ERNHART [2]

Tulane University

Many investigators (Reynolds, 1945; Sheffield, 1950; Stanley, 1952; Teichner, 1952; Wilson, Weiss, & Amsel, 1955) have reported faster extinction following a change in intertrial interval at the onset of extinction. This observation led to the hypothesis (Reynolds, 1945; Rothkopf, 1953, 1955; Teichner, 1952; Teichner & Holder, 1952) that the stimulus complex present at the beginning of a trial in a series of relatively massed trials includes temporally decaying traces of response-produced stimulation from the preceding trial or trials. The response, then, is conditioned to whatever magnitude of such traces remains at the end of the intertrial interval. A change in the interval changes this magnitude and, consequently, the stimulus complex to which the response is conditioned. According to a stimulus generalization interpretation, such a change results in a decrease in response strength. This study tests some of the implications of this analysis.

Within Hull's (1943) systematic formulation, a change in a stimulus complex decreases effective habit strength ($_S\bar{H}_R$) and, consequently, excitatory potential ($_SE_R$). When this stimulus change is due to a change in intertrial interval, it is also essential to consider reactive inhibition (I_R), which subtracts from $_SE_R$. A change from spaced to massed trials increases I_R; a change from massed to spaced trials decreases I_R. The confounding effects of I_R in this situation require a re-evaluation of the classical design for generalization experiments, in which it is customary to compare strengths of response to test and to training stimuli. It is implicitly assumed that the only factor differentiating performance on training and test stimuli is the training given to the former, i.e., *S*s trained equally to the two stimuli would respond equally to both. While generally valid, this assumption is not tenable when the stimuli under consideration are associated with other conditions which affect response strength, e.g., drive or reactive inhibition.

Stimulus generalization may also be evaluated by a comparison of test-trial performance with the performance of a control group trained only on the test-trial stimulus. Let A be a training stimulus and B a test stimulus. An experimental group, A-B, may be compared with two groups, A-A and B-B, trained and tested with A and with B, respectively. When the response strengths of A-A and B-B are identical, or nearly so, the test-trial decrement of A-B will be the same whether it is determined by comparison with A-A or with B-B.

These same habit strength relationships apply in the case of response-produced stimulation. Let A and B be, respectively, the stimulus traces

[1] This research represents the major portion of a dissertation submitted in partial fulfillment of the requirements for the PhD degree at Tulane University. The author is deeply indebted to Abram Amsel for his guidance during the planning, execution, and reporting of this study.

[2] Now with the Department of Psychiatry, Washington University School of Medicine.

present at the onsets of a spaced trial and of a massed trial. Although habit strengths are equal, the excitatory potential of A-A should be greater than for B-B because of greater I_R present with massed trials. The difference between A-B and an A-A group on test trials is compounded of stimulus generalization decrement and increased I_R due to the change from spaced to massed trials. A comparison of an A-B group and a B-B group, however, permits the independent evaluation of the effects of changes in ${}_S\bar{H}_R$, inasmuch as I_R has no differential effects on the test trials of these groups. A rest period for the dissipation of I_R prior to the test trials is, of course, necessary.

A design based upon A-B and B-B comparisons was employed to test several implications of the explanation that response decrements following changed intertrial interval are due to changes in stimulus components. (*a*) Systematic changes from a given interval to other intervals should produce response decrements which increase with the extent of change. (*b*) A varying interval, often used to avoid temporal conditioning, should reduce the specificity of the temporal aspects of the stimulus complex and reduce the decrements observed following a gross change in intertrial interval. (*c*) If a decrement following a change in intertrial interval is due to decreased ${}_S\bar{H}_R$, continued reinforcement should lead to greater increases in habit strength for experimental (changed interval) groups than for control groups. This assumes that learning is negatively accelerated.

METHOD

Subjects.—The *Ss* were 140 pigmented rats from two strains maintained at Newcomb College, Tulane University. They were divided into 14 groups, approximately equal in age (about 130 days), proportion of each strain, and sex. The *Ss* were gentled by handling.

Apparatus.—A startbox, 9 in. long, and a connected runway, 72 in. long, were 2½ in. wide and 3 in. high and were painted white. The runway opened into a black goalbox 13 in. long, 7 in. wide, and 4 in. high. The startbox, runway, and goalbox were separated, respectively, by white and black guillotine doors. Photoelectric units, 6 in. and 66 in. from the beginning of the runway, started and stopped a .01-sec. Standard Electric timer. In addition, breaking the second beam of light operated a motor-driven stainless steel loop which brought 0.1 ml. of water to the floor level of the goalbox. A separate Masonite retention box was 11½ in. square and 15 in. high with an opening of 8 sq. in. at the top. The entire room was well illuminated.

Deprivation and pretraining.—Twenty minutes before each day's session *S* was fed 4 gm. of dry Purina Laboratory Chow. One and a half hours after the beginning of the session, *S* was given 8 gm. of food with ad lib. water for 1 hr. and then placed on total deprivation for 22 hr., 10 min.

Each *S* was given eight days of 5-min. pretraining sessions occurring 5 min. before the time *S* would be run on experimental trials. For the first four days *S* was lifted in and out of the retention box, then for two days permitted nonreinforced exploration of the startbox, runway, and goalbox. For the last two days of pretraining, *S* was given six loops of water in the closed goalbox.

Experimental procedure.—Each *S* was given 21 reinforced trials per day for five days, three days on the training intertrial interval and two days on the test interval. The *S* was placed in the retention box for the appropriate interval, removed, and placed in the startbox. Two seconds later the doors were raised, and were lowered after passage of *S*. The *S* was removed from the goalbox 10 sec. after entry. Removal from the goalbox to the retention box took 2 sec., and moving *S* from the retention box to the startbox took 5 sec. For the balance of the intertrial interval, *S* remained in the retention box.

Experimental design.—In order to control for the possible effects of I_R, each experimental (changed interval) group was compared with a control group which had both training and test trials on the test-trial interval of the experimental group. Ten groups of *Ss* were required to establish the generalization gradients for changes from

massed to spaced and from spaced to massed trials. Four constant intertrial intervals were used: 10, 30, 60, and 120 sec. Of three experimental groups trained with the 10-sec. interval, one was changed to each longer interval (10–30, 10–60, and 10–120). Similarly, three groups were trained with the 120-sec. interval and tested at each smaller interval (120–60, 120–30, and 120–10). One of the four constant-interval control groups was trained and tested at each interval (10–10, 30–30, 60–60, and 120–120). These groups were controls for the experimental groups changed, respectively, to 10-, 30-, 60-, and 120-sec. intervals.

Four additional groups were trained with intervals varying in a predetermined order, two with intervals varying between 7 and 13 sec., averaging 10 sec., two with intervals varying between 84 and 156 sec., averaging 120 sec. One group from each of these was tested at 10 sec., the other at 120 sec. These groups are designated as 10V–10, 10V–120, 120V–10, and 120V–120.

Results

The data, comprising 105 running-time measures for each *S*, were transformed and condensed to a mean log running-time plus 1.00 for each block of seven trials. This provided 15 running-time scores for each *S*, one for each third of each day. These were employed in all analyses except those of first-trial decrement.

The differences among the four constant-interval training conditions on Days 1, 2, and 3 were not significant ($F = 1.48$, $df = 3/96$) in an analysis of repeated measures. A double classification analysis of the training trials of the 10–, 120–, 10V–, and 120V–sec. groups revealed no significant differences between varying and constant interval groups ($F = .46$), but did indicate that the massed groups ran more rapidly than the spaced ($F = 6.75$, $df = 1/116$, $P < .05$). This is a reversal of the usual observation of superior performance with spacing of trials.

Initial test-trial performance.—The data from the first postchange trial block (Trials 1–7, Day 4) were em-

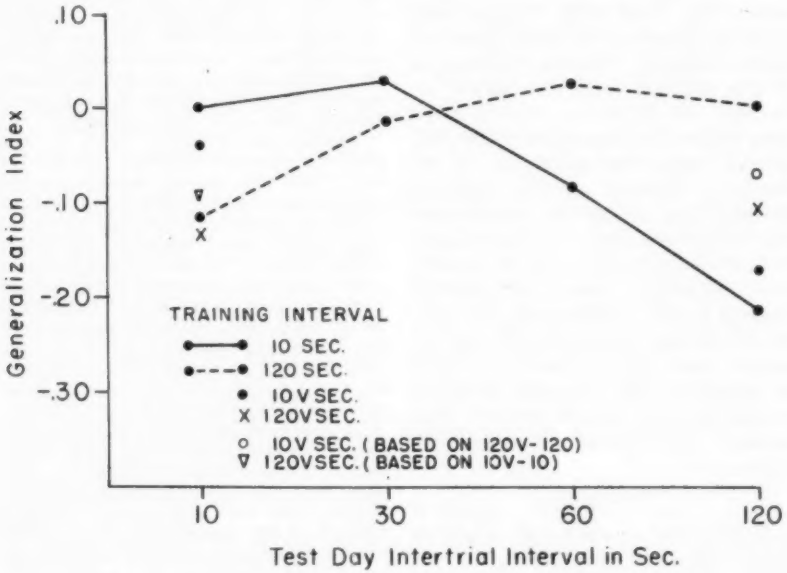

Fig. 1. Generalization gradients for Trials 1–7, Day 4.

ployed to plot the generalization gradients shown in Fig. 1. The generalization index was determined for each experimental group by subtracting its mean running-time score from that of its respective control group. The point on the ordinate marked "0" represents control group performance, i.e., points above this indicate more rapid running and points below this slower running on on the part of the experimental group. Points identified with varying training-intervals are discussed below.

The response decrements at the extreme points of each gradient were tested in an analysis of the 120–10 and 10–120 groups and their respective control groups (10–10 and 120–120). The difference between experimental and control groups was highly significant ($F = 13.14$, $df = 1/36$, $P < .001$). The two decrements, or differences between the 10–120 and 120–120 groups and between the 120–10 and 10–10 groups, did not differ significantly, as shown by the interaction term ($F = 1.15$, $df = 1/36$).

The gradients are similar in form, showing a slight improvement in performance following the smallest change and progressively increasing decrements with larger changes in intertrial interval. Each gradient was tested by a 2×3 analysis of variance of its three changed-interval groups and their respective control groups with experimental vs. control group as one classification and test-day interval as the other. Since each interval represents a different degree of change from the training interval, the interaction term reflects the increase in the difference between experimental and control groups from one test-trial interval to another. For the massed-spaced gradient experimental vs. control and the interaction terms were significant at the .05 level ($F = 5.37$, $df = 1/54$ and $F = 3.22$, $df = 2/54$, respectively). Corresponding tests for the spaced-massed gradient did not reach significance; however, the observation of greater decrement with the larger change in intertrial interval and the similarity of this gradient to the massed-spaced gradient militate against a hypothesis that the gradient is due to chance fluctuations.

The varying interval experimental groups, 10V–120 and 120V–10, were similar to the 10–120 and 120–10 groups in the amount of absolute decrement, differing, respectively, from the 120–120 and 10–10 groups at the .01 level ($F = 7.46$, $df = 1/36$). Since the treatment of the varying interval experimental groups differed from the constant interval control groups in two respects, gross change in interval size and change from a varying to a constant interval, comparisons were also made with Groups 10V–10 and 120V–120 which incorporated changes only from varying to constant intervals. Although differences were not significant, these groups ran slower than the corresponding constant interval control groups (10–10 and 120–120). Using these as control groups for the 120V–10 and 10V–120 experimental groups reduced the decrement scores for the latter. Decrement defined in this manner, controlling for a change from a varying to a constant interval, was not significant ($F = 1.40$). These relations are shown in Fig. 1. Although these decrements were less than for the corresponding constant interval groups, the differences between decrements failed to reach statistical significance ($F = 1.097$) as tested by the appropriate interaction term (experimental vs. control by varying vs. constant) of a three way analysis of variance (experimental vs. control by varying vs. constant by test-trial interval). In this analysis only the experimental vs. control

source of variance was significant ($F = 8.98$, $df = 1/72$, $P < .01$).

Persistence of response decrements.—The major relationships among the large-change groups (120–10, 10–120, 120V–10, and 10V–120) and their respective control groups (10–10, 120–120, 10V–10 and 120V–120) through the two days of reinforced practice on the postchange intertrial intervals are graphed in Fig. 2. These data are evaluated in the analysis of repeated measures summarized in Table 1. The overall decrease in running time from Day 4 to Day 5 did not reach statistical significance. Within each day, running time decreased from the first to the second trial block, then increased slightly on the third trial block. As in the training trial data, more rapid running was found with the 10-sec. interval than with the 120-sec. interval. This discrepancy between massed and spaced groups increased with trials on Day 4 and decreased on Day 5 (I × D × T interaction). There is no apparent explanation for this finding and it is probably attributable to chance.

The sources of variance of particular relevance to the hypotheses of this study are the E, the E × V, the E × D, and the E × T. As in the corresponding comparison for the first trial block, the decrement associated with a change in intertrial interval is highly significant. Inspection of

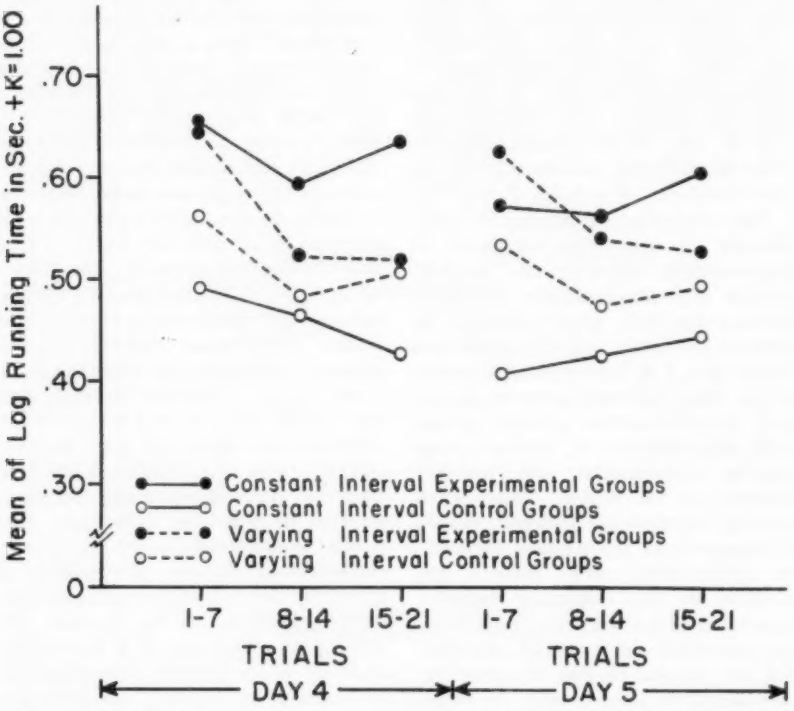

FIG. 2. Running-time scores for large-change groups for six-trial blocks on two test days.

TABLE 1

ANALYSIS OF VARIANCE OF RUNNING-TIME SCORES OVER TWO TEST DAYS

Source	*df*	*MS*	*F*
Between-*Ss*	(79)		
(E) Experimental vs. Control	1	1.2704	13.75**
(V) Varying vs. Constant Training Interval	1	.0091	—
(I) Test Interval	1	.5175	5.60*
E × V	1	.2061	2.23
E × I	1	.0806	—
V × I	1	.3422	3.71
E × V × I	1	.2385	2.58
Error (b)	72	.0924	
Within-*Ss*	(400)		
(D) Days	1	.0457	1.26
(T) Trial Blocks	2	.1431	8.64**
E × D	1	.0050	—
E × T	2	.0045	—
V × D	1	.0557	1.53
V × T	2	.0448	2.71
I × D	1	.0575	1.58
I × T	2	.0205	1.24
D × T	2	.0139	—
Higher Interactions[a]	26	—	—
$Error_1$ (w)	72	.0364	
$Error_2$ (w)	144	.0166	
$Error_3$ (w)	144	.0129	
Total	479		

* $P < .05$.
** $P < .001$.
[a] Of the 3-, 4-, and 5-way interactions, only I × D × T reached significance at the .05 level.

Fig. 2 shows that response decrement for the varying interval experimental groups, when based on varying interval control groups, continued to be less than for constant interval groups. The E × V interaction, testing this difference, failed to reach significance. The hypothesis that response decrement would decrease with further practice on the postchange interval received no support from the E × D or E × T interactions.

First-trial decrement.—First-trial decrement is defined as a decrease in response strength from the last prerest trial to the first postrest trial. These comparisons were made for Days 2 and 3 for the 60 experimental (changed interval) *Ss* and for Days 2 through 5 for the 40 control *Ss*. Data from the variable-interval groups were excluded. The percentages of *Ss* running slower on the first trial of a given day than on the last trial of the previous day were 73% ($N = 100$), 81% ($N = 100$), 73% ($N = 40$), and 90% ($N = 40$) for Days 2 through 5, respectively. Chi-square tests of the significance of the differences between these percentages and a chance level of .50 yielded P values of less than .01 for all but Day 4. First-trial decrement was not related to intertrial interval.

DISCUSSION

Decrements following changed intertrial intervals have been attributed to changes in the magnitude of response-produced stimulation. These changes alter the stimulus complex to which the response is conditioned and, consequently, reduce effective habit strength ($s\bar{H}_R$). According to this interpretation,

the amount of decrement associated with such changes should be positively related to the extent of the change in the interval. Data obtained by Rothkopf (1953) failed to demonstrate such systematic gradients. It is possible, however, that the six trials which he employed for training at each interval did not develop sufficient habit strength to the specific stimulus complex for differential decremental effects to be observed. In the present experiment, which employed 63 training trials, the obtained gradients clearly support the stimulus generalization interpretation.

Little attention has been given to the possible effects of a varying intertrial interval. It does not seem to have been employed systematically in experimental designs except as a means of avoiding temporal conditioning in classical conditioning procedures. That stimulus factors associated with the intertrial interval are not rendered irrelevant, or "noise," by the use of a varying intertrial interval is attested to by several features of the results reported herein. The training data revealed no apparent advantage of a fixed as opposed to a varying interval in acquisition. Of greater importance, perhaps, is the finding that absolute decrements following a change from a varying interval are significant and of the same order as decrements following a change from a constant interval.

The question then arises as to whether the decrement observed following the change from a varying interval is due to the gross change in the size of the interval or to the change from a varying to a constant interval. Mean running times of the varying-interval control groups fall between those of the corresponding constant-interval control groups and the varying-interval experiment groups. While these comparisons were not statistically significant, they suggest that the decrements obtained with varying-interval experimental groups are jointly produced by the large change in interval size and by the change from a varying to a constant interval. Controlling for the latter factor by use of the varying-interval control groups produces a consistently, although not significantly, smaller decrement for the varying-interval experimental groups and suggests that the generalization decrement associated with gross change alone is reduced when a varying interval has been employed in training.

The failure to find a greater rate of improvement, with continued reinforcement, for the experimental than for the control groups is surprising. Unfortunately, the form of the learning curve to a generalized stimulus does not appear to have been the subject of systematic thought or exploration. The prediction made in this experiment was based on the assumption that, to the extent that a stimulus complex is changed, learning to respond to the altered stimulus becomes more like naive or initial learning with its characteristic negatively accelerated development. The implications of the failure of this assumption, should it be demonstrated in other experiments, are obvious.

The first-trial decrement data provide further evidence that an extended rest period can lead to response decrements. Teichner and Holder (1952) have previously demonstrated the interaction of the facilitating effect of I_R dissipation and the decremental effect of the change in stimulation associated with the intertrial interval. Since the animals of present experiment were placed in the retention box for the appropriate interval before the first trial, the obtained first trial decrement appears to be more dependent upon response-produced stimulation than upon reinstatement of the immediate pretrial cues.

Summary

Four groups of 10 rats learned a running response in a straight alley with a 10-sec. intertrial interval. One group continued on this condition while three others were changed, respectively, to 30-, 60-, and 120-sec. intervals. Comparison of the latter three groups with control groups maintained on, respectively, 30-, 60-, and 120-sec. intervals indicated a significant increase in running time as a function of degree of change in the

interval. A corresponding portion of the experiment for changes from spaced (120-sec.) to relatively massed (60-, 30-, and 10-sec.) intertrial intervals yielded a similar, although not statistically significant, gradient of response decrement. Absolute decrements for additional groups changed from intervals varying about 10 and 120 sec. were of the same order as for groups trained with constant intervals; however, these decrements were reduced when the factor of a change from a varying to a constant interval was controlled. All decrements persisted undiminished through 42 trials of continued reinforcement. Performance on the first trial of each day was poorer than on the last trial of the previous day.

These results support the hypothesis that a change in intertrial interval changes the magnitude of response-produced stimulus traces from the preceding trial. This alteration of the stimulus complex to which the response is conditioned is assumed to decrease habit strength according to a stimulus generalization function.

REFERENCES

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

REYNOLDS, B. Extinction of trace conditioned responses as a function of the spacing of trials during acquisition and extinction series. *J. exp. Psychol.*, 1945, **35,** 81–95.

ROTHKOPF, E. Z. Distribution of practice and the temporal decay of response-produced stimuli. Unpublished doctoral dissertation, Univer. Connecticut, 1953.

ROTHKOPF, E. Z. Distribution of practice and the temporal decay of response-produced stimuli. *J. exp. Psychol.*, 1955, **49,** 33–38.

SHEFFIELD, V. F. Resistance to extinction as a function of the distribution of extinction trials. *J. exp. Psychol.*, 1950, **40,** 305–313.

STANLEY, W. C. Extinction as a function of the spacing of extinction trials. *J. exp. Psychol.*, 1952, **43,** 249–260.

TEICHNER, W. H. Experimental extinction as a function of the intertrial intervals during conditioning and extinction. *J. exp. Psychol.*, 1952, **44,** 170–178.

TEICHNER, W. H., & HOLDER, E. Reminiscence as a function of the amount of change in the intertrial interval. *J. exp. Psychol.*, 1952, **44,** 347–351.

WILSON, W., WEISS, E. J., & AMSEL, A. Two tests of the Sheffield hypothesis concerning resistance to extinction, partial reinforcement, and distribution of practice. *J. exp. Psychol.*, 1955, **50,** 51–60.

(Received September 16, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 4, 216–221

POSTSTIMULUS CUING IN IMMEDIATE MEMORY[1]

NANCY S. ANDERSON[2]

IBM Research Center

Several theories of human memory have hypothesized that memory preserves all of the original information which was presented, and that whatever error is made in recalling the information is a defect in the process of recall (Gomulicki, 1953). In an immediate memory task, *S* is asked to reproduce the complete message with each item in its correct position. If the number of items in a message is large, the act of writing the first items and the time involved may interfere with the recall of the remaining items.

In a recent study by Klemmer and Loftus (1958) an attempt was made to reduce the interference of writing on the recall of visual patterns. A class of 128 visual patterns were constructed of individually lighted line segments of a seven-element line matrix. If all lines of the matrix were illuminated a block figure eight was produced. Four matrices, each with a randomly chosen pattern of lines, were exposed tachistoscopically for 0.02-sec. to *S*s. After each exposure of the four patterns, a poststimulus cue was used to indicate to *S* to write only *one* of the four patterns. The percentage of correctly reproduced lines of a pattern was 14% higher for the poststimulus cued recall of patterns than when *S*s were required to write all four patterns. The authors conclude that the lower performance when writing four patterns was due to the interference of writing additional line elements. Similar cuing procedures were used by Augenstine, Blank, Quastler, and Wayner (1956) and Long and Garvey (1955). These studies also show that poststimulus cuing aids performance in the recall and identification of patterns.

The present studies were undertaken to compare immediate memory for spoken digits using a standard recall of all of the digits of a list and a poststimulus cue for recall of only part of the list of digits. In order to clarify the effects of the time and the interference involved in writing the digits, the two experiments also investigate recall as a function of a forced time delay between the last digit of the list and the cue for recall.

EXPERIMENT I

Procedure

Fourteen students from The King's College, Briarcliff Manor, New York, served as paid *S*s in this experiment.

Messages of 12 random numbers were presented to the group of 14 *S*s by a tape recording. The 12 digits were read in three groupings of 4 digits each, at an average rate of 2 digits/sec. For example: 4539 pause, 2164 pause, 3283. Each message of 12 digits and 2 pauses was 6 sec. long. All responses were recorded on prepared answer sheets. The poststimulus cue was presented by lighting combinations of three bulbs mounted linearly in a box clearly visible to all *S*s. The three lights corresponded to the three groups of digits read. Thus, if the first bulb was lighted, *S*s were to recall the first 4 digits; the second bulb, the second 4 digits; the third bulb, the last 4 digits. In addition, the combination of Lights 1 and 2, 2 and 3, 1 and 3 indicated the corresponding 8 digits. The combination of Lights 1, 2, and 3 indicated

[1] The author wishes to thank Lee Novick and Gregory Lockhead for their aid in data collection, and F. DeKalb who wrote the program for data reduction on the IBM 704.

[2] Now at the University of Maryland.

all 12 digits. All seven possible cuing combinations were employed: 3 sets of 4 digits, 3 sets of 8 digits, and all 12 digits.

The poststimulus cue for recall occurred at 0 sec. (immediately), 5 sec., or 30 sec. after the last digit of the message had been read. Each of the seven cued combinations occurred at random five times during each time delay. Each experimental session consisted of 35 messages for all cuing conditions at each of the three time delays (a total of 105 messages for each of 8 sessions). All seven combinations were presented at one time delay before testing under another time delay condition. The order of the three time delay conditions was varied at random over eight sessions. The *Ss* were informed of each time delay condition, but did not know what group(s) of digits were to be recalled until the poststimulus cue occurred.

The *Ss* were asked to write the 4, 8, or 12 digits on prepared answer sheets, containing sets of 12 squares. The 12 squares were clearly divided for 3 groupings of 4 each. The *Ss* were asked to record (in any temporal order they chose) each digit in the correct spatial location. If the cue indicated recall for the last 4 or 8 digits, *Ss* were permitted to record them in the first 4 or 8 response squares. Most *Ss* preferred to mark the 4 or 8 digits in the squares corresponding to their position in the message.

Results

The first two sessions were practice sessions. Only the data averaged over the last six sessions are presented here. Figure 1 shows the percentage correct of the digits as a function of cue delay averaged over all *Ss*. The curve for 4-number recall is the average percentage correct of the first, second, and third groups of digits. The curve for 8-number recall is the average percentage correct of the digits of Groups 1 and 2, 1 and 3, and 2 and 3. The curve for 12-number recall is the average percentage correct for the one condition requiring all 12 digits to be recalled. The results illustrate the effects of cue delay, number of digits cued, and the interaction of cue delay and number of digits cued. That is, the percentage

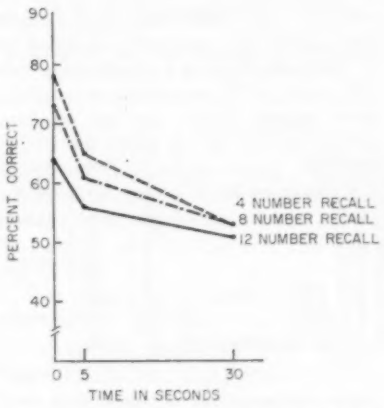

FIG. 1. Percentage correct for each recall condition as a function of 0, 5, and 30 sec. between the last digit and the cue for recall.

of correct scores of the 4- and 8-number recall were higher than the 12-number recall compared within each time delay. Inspection of the data showed this result for all 14 *Ss* at 0-sec. delay, 13 *Ss* at 5-sec. delay, and 11 *Ss* at 30-sec. delay. The percentage of correct recall for 4 numbers was higher than for 8 numbers at 0- and 5-sec. delays. This result was true for 11 of 14 *Ss*. The drop in performance due to increased delay was such that performance for 12-number recall at 0-sec. delay was higher than for 4- and 8- number recall at 30-sec. delay. Inspection of the data showed this latter result for all 14 *Ss*.

A four-variable analysis of variance, with *Ss* as one variable, was carried out and evaluated as a mixed model (McNemar, 1955). The data used in this analysis were the number of correct digits for each group of four digits, for each condition, for each *S*. The first-order variables of cue delay (D), group of digits (G), and number of digits (N) were significant. The following interactions were signifi-

TABLE 1

ANALYSIS OF VARIANCE OF THE NUMBER OF CORRECT OUT OF FOUR DIGITS

Source	*df*	*MS*	*F*
Ss	13	1.93	
Delay (D)	2	18.66	155.55*
Group (G)	2	31.32	13.10*
No. of Digits Cued (N)	2	3.37	84.25*
D × *Ss*	26	.12	
G × *Ss*	26	2.39	14.94*
N × *Ss*	26	.04	
G × D	4	.21	
D × N	4	.59	11.80*
N × G	4	.06	
D × N × G	8	.10	
D × N × *Ss*	52	.05	
D × G × *Ss*	52	.16	2.28*
N × G × *Ss*	52	.09	
Remainder (D × G × N × *Ss*)	104	.07	
Total	377		

Note.—The second-order interaction (G × *Ss*) is significant tested against either the D × G × *Ss* interaction or the residual error term.
* $P < .01$.

cant: (*a*) cue delay × number of digits cued, (*b*) group of digits × *Ss*, and (*c*) group of digits × cue delay × *Ss*.

Figure 2 shows the percentage of correct digits for the first, second, or third group of digits as a function of cue delay. The data for each group of four digits were averaged over all cue combinations which contained that group. For example, the scores for Position 1 were averaged for the first group of four digits when it occurred in the cued combinations of Groups 1, 1 and 2, 1 and 3, and 1, 2, and 3. It is clear that performance on the third group of digits was higher than performance on the second or first group of digits at each of the three time delays. This result was true for each of the individual cue combinations as well as the averages over the cue combinations. Although this was true for the average performance, the data for individual *Ss* showed that the two significant interactions with *Ss* were a result of higher performance at longer cue delays for some *Ss* on the first group of digits than on the third group of digits (interactions G × D × *Ss* and G × *Ss* as shown in Table 1).

One of the most striking results was the characteristic decrease in performance with an increase in cue delay for all conditions. It can be noted in Fig. 1 that the drop in performance from 0-sec. to 30-sec. delay was 25% for 4-digit recall, 20% for 8-digit recall, and 13% for 12-digit recall. Of this decrease over time, approximately one-half of it occurred between 0 and 5 sec. (13% for 4-digit, 12% for 8-digit, and 8% for 12-digit recall). The purpose of Exp. II was to investigate cued recall for time delays between 0 and 5 sec. in order to extend these results.

EXPERIMENT II

Procedure

Six students from The King's College served as paid *Ss* in this experiment. Only one *S* had

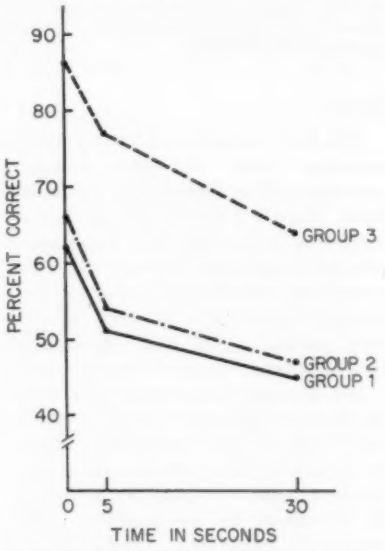

FIG. 2. Percentage correct as a function of the time between the last digit and the cue for recall. Each curve represents one of three Groups of four digits in the 12-digit series.

participated in the aforementioned experiment, although the other five *Ss* had participated in four previous memory experiments.

The procedure for message presentation and recording of digits was the same as in Exp. I. The only difference in procedure was that the poststimulus cue for recall occurred at 0 sec., 2 sec., or 4 sec. after the last digit of the message had been read. Each of the seven cued combinations occurred at random four times during each time delay for a total of 28 messages for all seven cuing combinations at each of three time delays. (A total of 84 messages for each of eight sessions.)

Results

The first three sessions were practice sessions and only the data averaged over the last five sessions are presented here. Figure 3 shows the percentage correct of the digits as a function of cue delay averaged over the six *Ss*. Again, the curves for 4-, 8-, and 12-number recall are averages of the first, second, and third groups of digits for each cue delay. These results illustrate the effects of Number of digits cued for 0-, 2-, and 4-sec. delays. Analysis of variance of the number of correct digits out of four (averaged over the three groups of digits) indicated that the variables of Number of digits, cue Delay, and the interaction of *Ss* × Number of digits were significant ($P < .01$). Percentage correct recall scores were higher for each of the six *Ss* for the 4-number recall than either the 8- or 12-number recall. However, the difference in performance for the 8- and 12-number recall was not the same for all *Ss*.

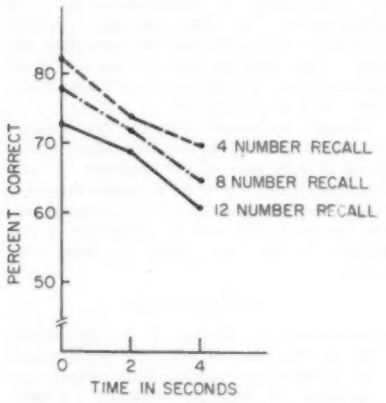

FIG. 3. Percentage correct for each recall condition as a function of 0, 2, and 4 sec. between the last digit and the cue for recall.

The decrease in performance as a function of cue delay was consistent for all six *Ss*. It should be noted that this decrease between 0 sec. and 4 sec. was approximately 12% and about half of that decrease occurred between 0 and 2 sec. (for each number of digits cued). Since the *Ss* from the two experiments were not the same, it was not possible to make comparisons between the 2-, 4-, and 5-sec. delay data. The average performance of the six *Ss* in Exp. II was higher at the 0-sec. delay than the average performance of the 14 *Ss* in Exp. I. This is most likely due to the fact that these six *Ss* had participated in several previous memory experiments.

The data describing performance for each of the three groups of digits were comparable to the results of Exp. I and are not presented here.

DISCUSSION

The results from both experiments indicated that poststimulus cuing for recall of only part of the message resulted in higher performance than recall of all the message. This supports the hypothesis that the act of writing the first digits of a message interferes with recall of the remaining digits in an immediate memory task.

Furthermore, the performance for all recall conditions decreased as a function of the delay between the end of the message and the cue for recall. The majority of the decrease in performance in Exp. I occurred between the 0- and

5-sec. delay in recall. The results from Exp. II showed that the decrease in performance from 0 to 4 sec. was regular with no apparent discontinuity. These results are in agreement with those of Klemmer and Loftus (1958) who also found that performance for cued partial recall of patterns decreased gradually for cue delays from 0.1 to 3.2 sec.

Reports from *S*s indicated that they rehearsed the digits between the presentation of the message and the cue for recall. During successive rehearsals *S* has no information about the accuracy of what he is rehearsing and is still uncertain about what he will be asked to recall. Both of these effects could account for the continuing decrease in performance up to 5 sec. The uncertainty of what to recall does not change throughout the time delay; however, *S* may change his strategy of rehearsing the digits. As the time delay increases, he may become more selective and rehearse only part of the message. The use of more than one behavioral strategy in an immediate memory task has been discussed by Pollack, Johnson, and Knaff (1959).

The results (Fig. 2) showing recall as a function of the position of the group of digits in the 12-digit series show the importance of the recency of the digits in recall. These results are in agreement with Deese (1957). There was no significant difference between the first and second group of digits. Thus, performance showed no enhanced effect due to the primacy of the digits. The absence of primacy effects in recall has been demonstrated by Raffel (1936). In her study on the recall of difficult nonsense material the absence of primacy occurred only when *S* was allowed to recall the syllables in any order he chose. In the present experiment, although *S* had to write the digits in the correct position in the series, he could record them in any temporal order he chose. The act of writing the last group of digits might have interfered with the recall of the first or second group of digits.

It must be noted that several *S*s did obtain higher recall scores on the first group of digits than on the last group of digits. During Exp. I, *S*s were asked to check which group of four digits was written first in the 8- or 12-digit series. These data showed that *S*s who wrote the first four digits first obtained higher recall scores on the first four digits. These results agree with Bousfield, Cohen, and Silva (1956) who developed an extension of Marbe's law showing that the rank order of recall of items was a negative function of the frequency of recall. The results of the present study indicate that the order of recall is influenced by the time delay, i.e., *S*s were more likely to write the first group of digits at the 5- and 30-sec. time delays. Since *S*s rehearse the digits, the additional rehearsal might enhance recall of the first group. Raffel (1936) found that instructions which allow rehearsal favor the recall of items in the beginning of the list.

Summary

The experiments investigated the effects of poststimulus cuing on recall of digits in an immediate memory task. Messages of 12 random numbers read in groups of 4 digits each were presented by tape recording. The *S*s wrote 4-, 8-, or 12 digits as indicated by the illumination of combinations of three cuing lights. The light cue was presented at various intervals after the last digit was read. The intervals were: 0 sec. (immediately), 5 sec., or 30 sec. in Exp. I and 0 sec., 2 sec., or 4 sec. in Exp. II.

Results indicated that (*a*) performance was highest when using a poststimulus cue for recall of only part of the message, (*b*) performance decreased with a forced delay between the presentation of the message and the cue for recall, and (*c*) performance was highest for the last group of digits read (recency).

References

Augenstine, L., Blank, A. A., Quastler, H., & Wayner, M. Human performance in information transmission. Flash recognition of familiar displays. *Univer. Illinois, Control Systems lab. Rep.*, 1956, No. R-69.

BOUSFIELD, W. A., COHEN, B. H., & SILVA, J. G. The extension of Marbe's law to the recall of stimulus-words. *Amer. J. Psychol.*, 1956, **69**, 429–433.

DEESE, J. Serial organization in the recall of disconnected items. *Psychol. Rep.*, 1957, **3**, 577–582.

GOMULICKI, B. R. The development and present status of the trace theory of memory. *Brit. J. Psychol.*, 1953, Monogr. Suppl. No. 29.

KLEMMER, E. T., & LOFTUS, J. P. Numerals, nonsense forms and information. *USAF Operat. Appl. Lab. tech. Rep.*, 1958, No. 57-2.

LONG, E. R., & GARVEY, W. D. The influence of sensory channel of cuing on the identification of aurally presented distorted words. *USAF WADC tech. Rep.*, 1955, No. 54–361.

MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1955.

POLLACK, I., JOHNSON, L. B., & KNAFF, P. R. Running memory span. *J. exp. Psychol.*, 1959, **57**, 137–146.

RAFFEL, G. Two determinants of the effect of primacy. *Amer. J. Psychol.*, 1936, **48**, 654–657.

(Received September 19, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 4, 222–234

THE IMMEDIATE RETENTION OF UNRELATED WORDS[1]

BENNET B. MURDOCK, JR.
University of Vermont

Many group learning curves from rote-learning studies are approximately exponential in form. The present study started with an attempt to find a rote-learning task such that the results of individual *S*s learning single lists would be a sufficiently close approximation to an exponential function so that this function could be used to describe the data. Describing the learning curves of individual *S*s by a mathematical function should not only make possible a more exact description of the data but also facilitate the discovery of interrelationships among variables.

It soon became apparent that in free-recall verbal learning the data of individual *S*s learning single lists could adequately be described by an exponential function. In free-recall verbal learning a list of items is presented one at a time and after each presentation *S* recalls as many items as possible. The words may be recalled in any order, and the order of presentation is randomly varied from trial to trial. The exponential function that describes the data is $R = c(1 - e^{-bn})$ where R is the number of items recalled after each presentation, c is an upper limit or asymptote, b is the rate constant or slope of the curve in a semilog plot, and n is the number of presentations. Thus, the exponential function shows how R varies as a function of n.

To avoid any confusion it is necessary to discuss briefly the relationship between learning and retention. The term "free-recall verbal learning" describes a type of rote-learning task in the same way that paired-associate learning or serial learning describe other types of rote-learning tasks. In free-recall verbal learning retention is measured after each presentation by the method of unaided recall. It is suggested that learning be considered the slope of the curve showing R as a function of n. Thus, learning would be the first derivative of R with respect to n, dR/dn. Such a distinction between retention and learning is analogous to the distinction between position and velocity in mechanics; just as velocity is the rate of change of position with respect to time so learning is the rate of change of retention with respect to the number of presentations. Such a definition of learning is of course consistent with the idea that is accepted by almost all psychologists; namely, that learning is an improvement in performance. This definition sidesteps all problems as to why the improvement occurs; in so doing it makes possible a clear separation between the phenomenon itself and the factors on which the phenomenon depends.

The fact that the curve $R = f(n)$ is exponential in form means that the slope of the line, or learning, is proportional to the number of items yet to be recalled; thus, $dR/dn = b(c - R)$. The learning becomes slower and slower with each successive presentation. Since learning is considered to be the slope of the retention curve (i.e., retention as a function of number of presentations) the obvious measure of learning is b, the slope of the straight line obtained by plotting $\ln[(c - R)/c]$ as a function of n, where ln is log to the base e. Thus, the rate constant b replaces such traditional measures of learning as number of trials to criterion or number of correct responses. Finally, it is assumed that the exponential describes the results of all *S*s in free-recall verbal learning; individual differences are assumed to be manifest in differing numerical values of the parameters b and/or c.

The present study reports a number of separate experiments, all of which dealt with the immediate retention of unrelated words. Some of the

[1] This article was facilitated by a Faculty Summer Research Fellowship from the University of Vermont. The author would like to thank Clinton D. Cook for invaluable advice at all stages of the project.

experiments studied retention after only one presentation, while others used repeated presentations. The experiments using repeated presentations were of course studies of free-recall verbal learning, but as has been pointed out, studies of free-recall verbal learning can be considered to be studying the phenomenon of retention. A number of the experiments were, either wholly or in part, designed to give evidence about practice effects. To allay any doubts as to the adequacy of the design used in many of the experiments it is necessary to state here that there are no practice effects, either warm-up or learning-how-to-learn. The preceding statement applies both to retention after one presentation and to free-recall verbal learning; evidence to support these assertions will be presented in the Results section.

The two experimental variables that were studied most intensively were the length of the list and the presentation time (i.e., time per item). Some of the experiments investigated the effects of length of list and presentation time upon retention following a single presentation; other experiments studied the effects of these two variables on free-recall verbal learning. For the latter experiments, since the effect of repeated presentations could be described by an exponentional function, the problem became one of determining the effects of length of list and presentation time upon the parameters b and c.

With two exceptions (Exp. IX and X) the learning material used in all the experiments were words drawn from the Thorndike-Lorge list of the 1,000 most common words in the English language (Thorndike & Lorge, 1944, pp. 267–270). As the words were essentially unrelated no attention was paid to the clustering of words during recall (Bousfield, 1953). Also, no attention was paid either to the rate at which the words were recalled during any given recall period (Bousfield & Sedgewick, 1944) or the order in which the words were recalled (Deese, 1957).

Procedure

After Exp. I a set of standardized conditions was established for the study of free-recall verbal learning. The standardized conditions (abbreviated SC) facilitated comparisons among different experiments, and also served as a point of departure for the study of the effects of the several experimental variables. Enough data were collected under SC to present both normative data and data on reliability. Also, in several experiments there was some departure from SC. Since these departures had no noticeable effects on free-recall verbal learning it was possible to consider them as irrelevant (Underwood, 1957, pp. 284–287).

The SC were as follows: The stimulus material consisted of a list of 30 words randomly selected from the Thorndike-Lorge list of the 1,000 most common words in the English language. The words were presented at a presentation time of 2 sec./word, and after each trial *S* wrote down on a mimeographed data sheet as many words as he could recall. The data sheets were arranged in two columns of 13 spaces each with the letters of the alphabet running down the page at the left of each column. Each word was to be written in the space by the letter with which it began. Four trials were given, and the periods of time allowed for recall were 90, 120, 150, and 150 sec. for Trials 1–4, respectively. The 30-word test list was preceded by one trial on a 10-word practice list with a 45-sec. recall period. Group testing was used, the words were printed freehand on 4 × 6-in. index cards, and the cards were shuffled between trials to approximate a random order.

The unequal recall times (i.e., 90, 120, 150, and 150 sec. for Trials 1–4, respectively) were used to make the group testing as comparable to individual testing as possible. When *S*s were tested individually (see Exp. V and VI) they were allowed as much time as they wanted for recall; it was found that they used progressively more time as the number of previous trials increased. The particular recall times selected for SC were

based on rough averages obtained from the individual testing.[2]

To determine if the unequal recall times for Trials 1–4 would affect retention, a group of 22 *Ss* was tested under SC. Then they were retested under identical conditions except that they were given the same recall time (150 sec.) for all four trials. For total number of words recalled the mean difference between the two conditions was less than one word, and the difference did not approach statistical significance ($t = 0.32$). Thus, it would appear that the unequal recall times used under SC had little effect on the number of words recalled.

Since the words were randomly selected from the Thorndike-Lorge list it was assumed that, except for sampling fluctuations, lists of the same length were always of comparable difficulty. Therefore, within a given experiment lists were not counterbalanced among conditions. However, it should be noted that most of the conclusions are based on the results of two or more different experiments, and different lists were used in different experiments.

As has been mentioned there are no practice effects, either warm-up or learning-how-to-learn. Therefore in some (though not all) of the experiments to be reported the order in which *Ss* received the various treatments was not counterbalanced. In general the primary consideration was simplicity of design; lists were never counterbalanced, and order was counterbalanced only when it could be done without increasing the complexity of the experimental design.

Finally, as also will be shown, there is no difference between visual and auditory presentation. Therefore, while visual presentation was used under the SC, auditory presentation was used instead in any experiment which required a presentation time of 1 sec./word or less. With auditory presentation all homonyms were deleted from the population of words before the samples were drawn.

In the following descriptions of specific experiments, it is to be understood that in all experiments making use of repeated presentations the procedural details were as described under SC unless specific mention to the contrary is made. In addition to the abbreviations already mentioned (R, c, b, n, and SC) the following abbreviations will also be used: L (length of list), PT (presentation time/item), t (total time required for presentation of the complete list), and R_1 (number of words recalled after one presentation). Further, in any numerical designation of a specific list the first number indicates L and the second number indicates PT; thus, 30–2 means a list of 30 words presented at a rate of 2 sec./word.

Experiment I.—The purpose was to determine if a learning-set developed. The *Ss* were tested under the SC on five successive sessions each spaced one week apart. A total of 29 *Ss* participated in one or more of the five sessions plus two make-up sessions, but only 19 *Ss* participated in all five regular sessions. In Exp. I and in all other experiments except where specifically noted the *Ss* were students of both sexes in the introductory psychology course who served as experimental *Ss* to fulfill a course requirement. No *S* served in more than one experiment.

Experiment II.—The purpose was to determine if a warm-up or fatigue effect occurred. All *Ss* ($N = 14$) learned six successive lists in one 2-hr. session. The first list was learned under the SC. Lists 1, 2, 4, and 5 were followed by a filler task consisting of 10 short-answer questions made up specifically for this purpose (e.g., "Who is the Governor of Vermont?"). The *Ss* were allowed 15 sec. to answer each question. List 3 was followed by an 8-min. rest period during which time *Ss* were free to leave the room.

Experiment III.—The purpose was to determine the effects of L on R_1. In Exp. IIIa the lists were 15-2, 30-2, 45-2, 60-2, and 75-2. A different group was used for each list, and each list was presented once. Auditory presentation was used. One 75-word list was recorded on a tape recorder and, for each group, the tape was simply stopped after the appropriate number of words had been read. There were 14, 55, 19, 29, and 25 *Ss* for the five groups, respectively. The 55 *Ss* who were given a 30-2 list were also given one trial on a second 30-2 list as soon as the first list was finished. In Exp. IIIb seven additional groups were tested, and each group was given one trial on each of two to four lists. The lists and the order in which they were presented were as follows: first group ($N = 10$): 400-1, 5-1, 5-2; second group ($N = 17$): 100-1, 200-2, 6-1, 6-2; third group ($N = 15$): 200-1, 100-2, 8-1, 8-2; fourth group ($N = 24$): 45-1, 150-1, 10-1, 10-2; fifth group ($N = 30$): 75-1, 15-1, 30-1; sixth group ($N = 9$): 60-1, 60-2, 30-2;

[2] Although the two experiments using individual testing are described as Exp. V and VI, they were actually the first experiments to be performed. The numbering of the experiments is based on expository convenience, not chronology.

seventh group ($N = 13$): 60-1, 150-2. Thus, for Exp. IIIa and IIIb combined, there were at a *PT* of 2 sec./word Lists 5, 6, 8, 10, 15, 30, 45, 60, 75, 100, 150, and 200 words in length. For a *PT* of 1 sec./word there were all these lengths plus a 400-word list. In all a total of 260 *Ss* participated in Exp. III.

Experiment IV.—The purpose was to determine the effects of *L* on R_1 using the method of whole presentation. Three lists (25, 50, and 75 words long) were mimeographed. Three different groups of 18 *Ss* each were used. Each group was given one trial on each list, and the order in which the three groups learned the three lists was counterbalanced. For each list *Ss* were given 30 sec. to study the list in any way they wished, and then they were tested for recall.

Experiment V.—The purpose was to determine if the results of individual *Ss* learning single lists in free-recall verbal learning are adequately described by an exponential function. There were 18 *Ss* tested individually on a 30-2 list. Both practice and test list were carried to a criterion of one perfect trial except that the experimental session was limited to 1 hr.; if *S* did not reach criterion on the test list within the allotted time the experiment was terminated. Written recall was not used; *S* said aloud all the words he could recall and *E* recorded them on a check list.

Experiment VI.—The purpose was to determine the effect of *L* on *b* and *c*. One *S* was tested individually for 22 sessions, and at each session a different list was learned. There were six lists of 20-2, four lists each of 30-2, 40-2, and 50-2, and two lists each of 75-2 and 100-2. The order in which these 22 lists were learned was approximately counterbalanced. Both practice and test list were learned to a criterion of one perfect trial. As in Exp. V, *E* recorded the words recalled on a check list. The *S* was an undergraduate psychology major who volunteered for the experiment, and she was paid for her services.

Experiment VII.—The purpose was to determine the effect of *L* on *b* and, in part, to replicate Exp. VI with additional *Ss*. This experiment consisted of four sessions. On each session there was one trial of a 10-2 practice list, four trials of a test list, the filler task described under Exp. II, and then four trials of a second test list. The eight test lists were 15-2, 20-2, 25-2, 30-2, 35-2, 40-2, 50-2, and 60-2. The order was counterbalanced in that each session one "long" (35 words or more) and one "short" (30 words or less) list was used, and the order of long and short alternated on successive sessions. The first test list learned the first day was 30-2. There were 22 *Ss* in Exp. VII.

Experiment VIII.—The purpose was to determine the effect of *PT* on R_1 and *b*. The procedure was identical with that of Exp. VII except that the experimental variable was *PT* and not *L*. The eight lists were 30-$\frac{1}{2}$, 30-$\frac{3}{4}$, 30-1, 30-1$\frac{1}{4}$, 30-1$\frac{1}{2}$, 30-2, 30-3, and 30-4. The "fast" (1$\frac{1}{2}$ sec./word or less) and "slow" (1$\frac{1}{2}$ sec./word or more) rates were counterbalanced as in Exp. VII, and the first test list learned the first day was 30-2. Auditory presentation by a tape recorder was used. A total of 18 *Ss* participated in one or more sessions, but only 13 *Ss* participated in all four sessions.

Experiment IX.—The purpose was to determine the effect of frequency of usage on R_1 and *b*. The procedure was identical with that of Exp. VII except that the experimental variable was frequency of usage and not *L*. Frequency of usage was varied by dividing the 30,000 words of the Thorndike-Lorge list into eight subgroups, each subgroup containing approximately the same number of words. The frequency of usage of the eight subgroups were 100+/million, 50–99/million, 10–49/million, 4–9/million, 2–3/million, 1/million, 5–17/18 million, and 4/18 million. The "high" (4–9/million or more) and the "low" (2–3/million or less) groups were counterbalanced as in Exp. VII, and the 100+/million (comparable to the 1,000 most common words) was the first test list learned the first day. All lists were 30-2. A total of 22 *Ss* participated in one or more sessions, but only 20 *Ss* participated in all four sessions.

Experiment X.—The purpose was to determine the effect of emotional value on R_1 and *b*.[3] Each of 50 *Ss* (friends and sorority sisters of *E*) rated 30 words from Stagner's list (Stagner, 1933) on a 7-point scale of emotional value. For each *S* two lists were constructed, one of the 12 highest-rated and one of the 12 lowest-rated words. Each *S* then learned both lists to a criterion of one perfect trial. The *Ss* were tested individually, and the order in which the two lists was learned was counterbalanced. No practice list was used, *E* recorded the words that *S* was able to recall, and *PT* was 3 sec./word.

Experiment XI.—The purpose was to replicate Exp. II, to determine R_1 as a joint function of *L* and *PT*, and to study the effects on R_1 and *b* of variations in *L* and *PT* holding

[3] This experiment was designed and conducted by Cynthia Marvin in a course in experimental psychology taught by the author.

total presentation time (t) constant. In Exp. XIa (Session 1) the first half of Exp. II was replicated exactly except that the order of the lists was changed from 1, 2, 3, to 3, 1, 2. In Exp. XIb (Session 2) following a practice list R_1 was determined for 18 different lists. The lists were of three different lengths (15, 45, and 75 words) and six different presentation times ($\frac{1}{2}$, $\frac{3}{4}$, 1, $1\frac{1}{2}$, 2, and 3 sec./word). The order in which the lists were presented was determined by a table of random numbers. Auditory presentation was used, and there was a 2-min. rest after the sixth and after the twelfth list. In Exp. XIc (Session 3) there was a practice list, a 60-1 list, a filler task, a 20-3 list, a filler task, and a 40-$1\frac{1}{2}$ list. Auditory presentation was used; for all three sessions $N = 16$.

Experiment XII.—The purpose was to replicate Exp. II and XIa, and to get further information on the effects of L on R_1. In Exp. XIIa (Session 1), following a practice list, *Ss* were tested for R_1 after four lists of 15-2, 30-2, 45-2, and 60-2 each. The order of these 16 lists was counterbalanced in blocks of four, and auditory presentation was used. In Exp. XIIb (Session 2), Exp. II and XIa were replicated exactly except that completely different lists were used. For the two sessions the *Ns* were 16 and 17, respectively.

Results and Discussion

Instead of presenting the results of each experiment separately, the conclusions that have been drawn will be stated and, following each conclusion, its evidence. This method of presentation will be used because in most cases each conclusion is based on more than one experiment. In some cases corroborative evidence from other studies in the literature is available; such evidence will be cited where appropriate.

Conclusion 1.—The numerical value of the asymptote c is the number of words in the list. Evidence: (*a*) For each of the 22 lists learned by *S* in Exp. VI a least-squares solution of the exponential was used which gave a numerical value for both b and c. Specifically, the least-squares solution was used on the approximation $\Delta R/\Delta n = bc - bR$ where ΔR was the upper envelope of the curve and R was the midpoint of each ΔR interval. The mean values for c were 21.0, 30.5, 40.5, 51.2, 80.2, and 101.3 words for the 20, 30, 40, 50, 75, and 100-word lists, respectively. The coefficient of determination (r^2) was 99.7%; thus, practically all the variability in c can be attributed to L. (*b*) The same least-squares solution was used on the data of each of the 18 *Ss* in Exp. V. For all *Ss* the mean value of c was 31.0 words, and this did not differ significantly from the expected value of 30 words ($t = 1.03$, $P > .05$). (*c*) Bruner, Miller, and Zimmerman (1955) used the method of free-recall verbal learning with four lists of 8, 16, 32, or 64 words. Their recall data are plotted in percentages in the top half of their Fig. 1 (p. 188). All groups reached a criterion of at least 80% correct, and their curves suggest that, with a sufficient number of trials, all groups would reach or come close to 100% recall.

The fact that the asymptote c is the same as L means that one of the parameters in the exponential equation can be identified with an experimental variable which is under the control of *E*. Also, it greatly simplifies the work of fitting the exponential to the data. Knowing the value of the length, L, of a list, substituting L for c, and writing the equation in the logarithmic form gives the equation $\ln[(L - R)/L] = -bn$. Thus, a plot of $\ln[(L - R)/L]$ against n gives a straight line with a slope of b and an intercept of zero.

A standard method has been used to solve for b. For SC there are actually five points to determine the curve. These are the number of words recalled after 0, 1, 2, 3, and 4 trials (when $n = 0$ then $R = 0$). It is necessary to use the least-squares formula for a straight line that passes through the intersection of the two axes (Bennett & Franklin, 1954, p. 232).

A least-squares solution for the exponential in log form does not minimize the sums of squares of deviations for the exponential form. In fact the log form introduces a systematic bias, weighting the last trial most heavily and the 0th trial least heavily (Murdock & Cook, 1960). In an attempt to correct for this bias, Trials 1, 2, 3, and 4 were weighted by a factor of 4, 3, 2, and 1, respectively. Although this weighting is arbitrary and does not result in a true least-squares solution for the exponential it is at least objective and simple to use; also, the data simply do not warrant a more precise treatment.

It is to be understood, then, that in all the analyses to follow, the parameter c in the exponential is assumed to have the numerical value of L. Furthermore, in all the analyses to follow the numerical value of b was determined by the weighted least-squares method described in the preceding paragraph. Finally, in all cases a separate numerical value of b was always calculated for each list learned by each S. Thus, when mean values are presented, they will be mean b values and not the b obtained from the mean trial-by-trial values.

Conclusion 2.—The data for individual *S*s learning single lists is adequately described by an exponential function. Evidence: (*a*) There were a total of 101 *S*s who learned under SC: 29 *S*s from Exp. I, 14 *S*s from Exp. II, 22 *S*s from Exp. VII, 20 *S*s from Exp. IX, and 16 *S*s from Exp. XI. In all cases *S*s were given four trials on a 30-2 list and, except for the practice list, it was the first list they had learned. After the value of b had been determined for each S a standard error of estimate, ($\sigma_{R \cdot n}$) was computed for each S. The distribution of $\sigma_{R \cdot n}$ for the 101 *S*s was positively skewed with a median of 1.4 words and a semi-interquartile range of 0.5 words. Thus, on the average, two-thirds of the obtained values did not differ by more than 1.4 words from the best-fitting exponential. (*b*) For the 18 *S*s of Exp. V, b was determined in the same manner (i.e., first four trials only), but the $\sigma_{R \cdot n}$ was computed from the total number of trials given each S. The median $\sigma_{R \cdot n}$ was 1.7 words and the semi-interquartile range was 0.4 words.

As no mathematical function ever describes experimental data perfectly, the question as to whether a given set of data is "adequately" described by a particular function is essentially a matter of judgment. In the present case the standard error of estimate has been used as a measure of goodness of fit. As has been shown, the deviations between the obtained data and the mathematical function average approximately 1.5 words; this would seem to be an adequate fit.

Another way of testing the appropriateness of a given function is to compare its goodness of fit with that of other possible functions. In the present case the most reasonable comparison is with the stochastic model of Bush and Mosteller (1955) which has also been suggested for free-recall verbal learning. For the Bush and Mosteller model it is necessary that *S*s reach a criterion of one perfect trial in order to obtain an estimate of α_1. There were 13 *S*s in Exp. V who reached criterion within the allotted time, and for each S the necessary parameters of the Bush and Mosteller model were determined. The value of p_0 was determined by the ratio $(N - 1)/(N_0 - 1)$ and, for α_1 $\bar{T}_2$ was the mean number of nonrecalls. As a measure of goodness of fit for these 13 *S*s, the median $\sigma_{R \cdot n}$ for the Bush and Mosteller model was 2.2 words whereas for the exponential function the median $\sigma_{R \cdot n}$ was 1.6 words. A t test showed that the exponential function gave a significantly smaller $\sigma_{R \cdot n}$ than did the stochastic model ($t = 4.71$, $P < .01$). Therefore, insofar as it was possible to make a comparative test, it would seem that the exponential provides a more accurate

description of free-recall verbal learning than does the Bush and Mosteller stochastic model.

Conclusion 3.—There is no difference between group testing with written recall and individual testing with oral recall. Evidence: (*a*) In Exp. V there were 18 *Ss* who were tested individually with oral recall. The mean value of *b* for these 18 *Ss* was 0.41, and this value did not differ significantly from the mean *b* value of 0.39 obtained for the 101 *Ss* tested under SC ($t = 0.64$, $P > .05$).

Conclusion 4.—There is no difference between auditory presentation and visual presentation. Evidence: (*a*) In Exp. VIII there were 18 *Ss* who were tested under SC except that auditory presentation replaced visual presentation. The mean value of *b* for these 18 *Ss* was 0.37, and this value did not differ significantly from the mean *b* value of 0.39 obtained for the 101 *Ss* tested under SC ($t = 0.55$, $P > .05$).

The fact that individual testing with oral recall and auditory presentation do not affect the results makes it possible to combine the 18 *Ss* of Exp V and the 18 *Ss* of Exp. VIII with the 101 *Ss* tested under *SC*. Since there is no learning-how-to-learn effect, and since their results do not differ significantly from the 101 *Ss* of SC ($b = 0.38$, $t = 0.13$, $P > .05$), the 17 *Ss* of Exp. XIIb can also be added. Thus, there were a total of 154 *Ss* who learned a 30-2 list either under SC or under conditions which were equivalent to SC; the data of these 154 *Ss* can be used to obtain norms for free-recall verbal learning.

Conclusion 5.—The distribution of the rate constant *b* is positively skewed, but its distribution is normalized by a logarithmic transformation. Evidence: (*a*) A histogram giving the distribution of *b* for all 154 *Ss* learning a 30-2 list is shown in Fig. 1

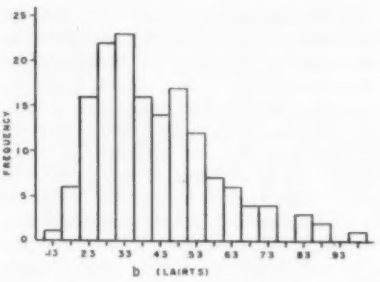

FIG. 1. Histogram showing distribution of *b* for 154 *Ss*.

and, as can be seen, the distribution is positively skewed. This skewness is statistically significant ($t = 4.73$, $P < .001$) when tested by a method suggested by Snedecor (1956, pp. 199–200). (*b*) A logarithmic transformation of *b* results in a distribution that does not differ significantly from a normal distribution ($\chi^2 = 4.10$, $df = 6$, $P > .05$).

For all 154 *Ss* the mean value of the rate constant (obtained by taking the antilog of the mean of the transformed scores) was 0.39, and one *SD* covered the range from 0.26 to 0.57. As an indication of the stability of *b*, the 154 *Ss* came from eight different experiments and these different experiments were considered as separate samples. An analysis showed that with samples of the size used (approximately 15 to 20 *Ss*) two-thirds of the samples did not vary by more than 7% from the total mean of 0.39.

To discuss the nature of the rate constant more completely, it should be pointed out that *b* is a measure of the rate at which the curve approaches the asymptote *c* (*L*). The unit of measure of *b* is reciprocal trials; we would suggest the term "lairt." The reciprocal of the rate constant indicates the number of trials required to attain approximately 63% of *L*. Thus, to say that the total mean was 0.39 lairts indicates that, on the average, *Ss* required 2.6 trials to reach a criterion of 19 out of 30 words

TABLE 1

MEANS AND *SD*s OF LOG 10*b* FOR EXP. I

Session	Group A $N = 19$		Group B $N = 9$	
	Mean	*SD*	Mean	*SD*
1	.65	.16	.57	.19
2	.58	.17	.60	.21
3	.63	.16	.64	.20
4	.65	.16	.60	.16
5	.60	.14		

(i.e., 63% of *L*). Actually, *b* is like *SD* in that it can be used to specify any desired criterion of mastery. Thus, in 0.5/*b* trials *S* should reach 39% *L*, in 2/*b* trials 86% *L*, in 3/*b* trials 95% *L*, etc.

Because the distribution of *b* was positively skewed all statistical tests (including those already presented under Conclusions 3 and 4) are based on a logarithmic transformation of *b*. When means and *SD*s of *b* are reported it is to be understood that these values are the antilogs of the transformed scores.

Conclusion 6.—There is no learning-how-to-learn effect in free-recall verbal learning. Evidence: (*a*) In Exp. I there were 19 *S*s (Group A) who participated in all five regular sessions. An analysis of variance showed that the effect of sessions was not statistically significant ($F = 2.01$, $P > .05$). The means and *SD*s of log 10*b* are shown in the upper part of Table 1 for each of the five sessions. (*b*) In Exp. I there were an additional nine *S*s (Group B) who participated in at least four different sessions. An analysis of variance showed that with these *S*s the effect of sessions was not statistically significant ($F < 1.00$). The means and *SD*s of log 10*b* are shown in the lower part of Table 1 for each of these four sessions. (*c*) As has already been indicated, the mean value of *b* for Session 2 of Exp. XII (i.e., Exp. XIIb) did not differ significantly from the mean value of *b* for SC.

Greenberg and Underwood (1950) found that recall following a 10-min. retention interval did not improve with practice. The present results are consistent with these findings; the slope of the retention curve did not vary systematically as a function of stage of practice.

Conclusion 7.—There is no warm-up effect in free-recall verbal learning. Evidence: (*a*) An analysis of variance of the results of Exp. II showed that the effect of lists *was* statistically significant ($F = 7.14$, $P < .001$). The means and *SD*s of log 10*b* are shown in Table 2 and of the six lists learned only List 2, according to Tukey's tests (Edwards, 1954, pp. 330–335) differed significantly from the rest. (*b*) Experiment XIa was an exact replication of the first half of Exp. II except that the order of the lists was 3, 1, and 2 instead of 1, 2, and 3. Again the effect of lists was statistically significant ($F = 14.16$, $P < .001$) and again, according to Tukey's tests, only List 2 differed significantly from the rest. However, in Exp. XIa List 2 was the third list learned, and for the second list learned (i.e., List 1) the mean value of *b* was actually lower (though not significantly so) than the first list learned. Thus it would appear that the significant *F*s resulted from a particular list which was easier to learn and not from any

TABLE 2

MEANS AND *SD*s OF LOG 10*b* FOR EXP. III ($N = 14$)

Measure	Session					
	1	2	3	4	5	6
Mean	.53	.66	.55	.55	.54	.49
SD	.15	.19	.21	.19	.20	.17

warm-up effect. (*c*) As a further check, Exp. XIIb was an additional replication of the first half of Exp. II except that three completely different 30-2 lists were used. An analysis of variance showed that the effect of lists was not statistically significant ($F < 1.00$). (*d*) In Exp. X a number of *S*s reached criterion so quickly that it was not possible to calculate a numerical value for b. Instead, trials to criterion was used as the measure of learning and the mean difference between the first and second list was a nonsignificant 0.3 trials ($t = 0.83$, $P > .05$).

With the large number of different lists used it was inevitable that some lists would be appreciably harder or easier than average. The List 2 of Exp. II seemed to be the only list of all those used that was sufficiently deviant so as to result in statistically significant differences in learning.

Conclusion 8.—The reliability of b is .77. Evidence: (*a*) The correlation coefficient of .77 is based on 76 *S*s: 29 *S*s from Exp. I, 14 *S*s from Exp. II, 16 *S*s from Exp. XIa, and 17 *S*s from Exp. XIIb. The correlation was based on the first two 30-2 lists learned by each *S* except for the 14 *S*s of Exp. II. In Exp. II the second list was the deviant list; therefore, the third 30-2 list was substituted for the second 30-2 list.

Conclusion 9.—The relationship between the rate constant b and frequency of usage (FU) as measured by the Thorndike-Lorge word count is given by the formula, $b = .048 \log FU + 0.31$. Evidence: For the data of Exp. IX the coefficient of determination (r^2) was 66%, and the slope of the line differed significantly from zero ($t = 3.43$, $df = 6$, $P < .05$).

The linear relationship between b and log FU suggests that the same relationship which holds for perceptual recognition thresholds (Howes & Solomon, 1951) also holds for learning. The results are also in essential agreement with those of Hall (1954) for free-recall verbal learning.

Conclusion 10.—There is no warm-up effect for R_1, the number of items recalled after one presentation. Evidence: (*a*) There were 55 *S*s in Exp. IIIa who were given one trial on each of two successive 30-2 lists. The mean R_1 difference was -0.3 words, and this difference was not significant ($t = 0.76$, $P > .05$). (*b*) In Exp. XIIa, *S*s were given one trial on each of 16 lists. These 16 lists were divided into four quarters and the differences among the four quarters were not significant ($F = 1.44$, $P > .05$). (*c*) An analysis of variance of R_1 for the three lists of Exp. XIIb showed that the differences among lists were not significant ($F = 1.14$, $P > .05$).

The data of Exp. XIb cannot be used to test for warm-up because the order in which the lists were presented was randomized, not counterbalanced. However, in addition to the above evidence there is an unpublished experiment by A. J. Babick and the author which is relevant. This experiment tested R_1 for 18 successive 25-1 lists; each of 18 *S*s was tested individually. The 18 lists were divided into six groups of three lists each, and an analysis of variance showed that the differences among lists were not significant ($F = 2.24$, $P > .05$).

Conclusion 11.—$R_1 = kt + m$ where $t = L \times PT$. Evidence: The evidence for Conclusion 11 is shown in Table 3 which gives the values of r^2, k, and m for the six relevant experiments (considering Exp. IIIa and III as separate experiments even though five points from the former are included in the 25 points of the latter). One source of evidence for a

TABLE 3

VALUES OF r^2, k, AND m FOR $R_1 = kt + m$

Exp.	Ind. Var.	Range	N	r^2	k	m
IIIa	L	15–75 words	142	98%	.071	5.9
III	L	5–400 words	260	88%	.056	5.8
VII	L	15–60 words	22	72%	.063	6.3
VIII	PT	½–4 sec.	13	92%	.080	4.8
XIb	PT	½–3 sec.	16	98%	.025	6.3
XIIa	L	15–60 words	16	90%	.028	6.4
Median				91%	.060	6.1

linear relationship is the coefficient of determination, r^2; the median r^2 for the six experiments was 91%. Perhaps even more important it should be noted that the various experiments, despite different lists, different *S*s, different methodologies, different conditions, and even different experimental variables, all gave numerical values for k and m that were of the same order of magnitude.

It appears that the effects of L and PT can be subsumed under the total time, t. Thus, t is the critical variable and it is a multiplicative function of L and PT. The numerical values for the entries of k and m in Table 3 were obtained by making use of this multiplicative relationship irrespective of whether L or PT was the variable actually manipulated in the experiment itself.

Since the slope k is positive, R_1 increases either as PT increases or L increases, the other held constant. There is considerable supporting evidence from Woodworth (1938, pp. 19–20) for the general finding that the number of items recalled increases as the list gets longer. Nonetheless, it was still rather surprising to find, for instance, that while the average recall of a 400-word list was 35 words no *S* was able to recall more than 10 words out of a 15-word list. Incidentally, it should be pointed out that, according to our data, it is misleading to emphasize the fact that the percentage of words recalled decreases as L increases. Actually, the percentage of words recalled over and above m stays constant; specifically, $(R_1 - m)/L = kPT$.

Since t is the critical variable, it should be possible to vary L and PT simultaneously while keeping t constant and obtain the same recall. In Exp. XI there was a 20-3, a 30-2, a 40-1½, and a 60-1 list. Since for all four lists t was the same (60 sec.) R_1 should therefore be the same for all four lists. The mean values of R_1 for the four lists were 9.3 words, 9.3 words, 9.6 words, and 8.4 words, respectively; the differences were not significant ($F < 1.00$).

It is suggested, then, that more words are recalled from a longer list because the longer list requires more time to present. If this is so, R_1 for lists of different lengths presented for the same length of time by the method of whole presentation should not differ since t is the same. Experiment IV was conducted to test this and, as predicted, length was not a significant variable ($F = 2.48$, $P > .05$).

Using the median values of k and m shown at the bottom of their respective columns in Table 3 the formula becomes: $R_1 = .06t + 6.1$, where t is measured in seconds. Assuming comparable samples it should be possible to predict results from other studies. Peters (1936) used a 10-1 list; $\bar{R}_1 = 6.7$ words, obtained $\bar{R}_1 = 5.2$ words (Table 2, p. 575). Bousfield, Sedgewick, and Cohen (1954) used a 60-3 list; $\bar{R}_1 = 16.9$ words, obtained R_1 at the end of 2 min. = 16.2 words. Deese (1957) used a 20-1 list; $\bar{R}_1 = 7.3$ words, obtained $\bar{R}_1 = 5.7$ words. Deese and Kaufman (1957) used a 32-1 and a 10-1 list; $\bar{R}_1 = 8.0$ words and 6.7 words, and as best could be determined from their Fig. 1 (p. 182) obtained $R_1 = 8.5$ words and 5.8 words, respectively. Unfortunately, several other studies which have tested R_1 have either not used a fixed PT or else have not reported it. While the above predictions are fairly accurate, on the average they tend to be a little high. Most of these experiments selected the words from the 30,000 most common words, so the

obtained values should be a little lower than would be predicted on the basis of the 1,000 most common words.

Given the linear relationship between R_i and t and the exponential function for free-recall verbal learning, it is possible to state the relationship between b, L, and PT. This relationship can be written:

$$e^{-b} = 1 - kPT - \frac{m}{L} \qquad [1]$$

To illustrate how Equation 1 works, for a 30-2 list $L = 30$ words and $PT = 2$ sec./word. Taking median values so $m = 6.1$ words and $k = .06$ words/sec. then $e^{-b} = .677$, or $b = .39$ lairts. The predicted value agrees perfectly with the mean value for b obtained under SC, even though four of the six experiments used to determine the median values which entered into the determination of $\tilde{b}$ were completely different from those experiments which entered into the determination of b.

In Exp. VII, L was systematically varied, and this experiment can be used to test Equation 1. In Exp. VII, $PT = 2$ sec./word; assuming the median values for k and m Equation 1 becomes:

$$\tilde{b} = -\ln\left[.88 - \frac{6.1}{L}\right] \qquad [1a]$$

The symbol "ln" represents log to the base e. The predicted and obtained

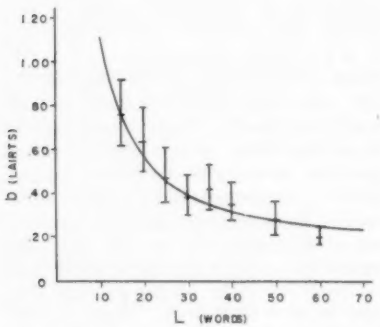

FIG. 2. Predicted and obtained results for b as a function of L.

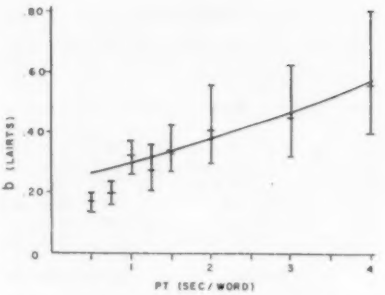

FIG. 3. Predicted and obtained results for b as a function of PT.

results are shown in Fig. 2. The smooth curve gives the predicted values and midpoints of the vertical lines are the obtained group means. The extent of the vertical lines indicates the 99% confidence interval. All points show satisfactory agreement.

In Exp. VIII, PT was systematically varied, and this experiment can also be used to test Equation 1. In Exp. VIII, $L = 30$ words; assuming the median values for k and m Equation 1 becomes:

$$\tilde{b} = -\ln\left[.80 - .06\,PT\right] \qquad [1b]$$

The predicted and obtained results are shown in Fig. 3 and, except for the two fastest presentation times, all points show satisfactory agreement.

Finally, in Exp. XI there were four lists in which L and PT were simultaneously varied but t was constant. With $t = 60$ sec. and again using the median values Equation 1 becomes:

$$\tilde{b} = -\ln\left[1 - \frac{9.7}{L}\right] \qquad [1c]$$

The predicted and obtained values are shown in Fig. 4 and, except for the fastest presentation time (i.e., the 60-1 list) all points show satisfactory agreement. Considering the results as a whole, they seem to provide fairly convincing evidence in support of Equation 1.[4]

[4] Analyses of variance of Exp. VII, VIII, and XI showed that, in each case, group differences were significant beyond the .001 level.

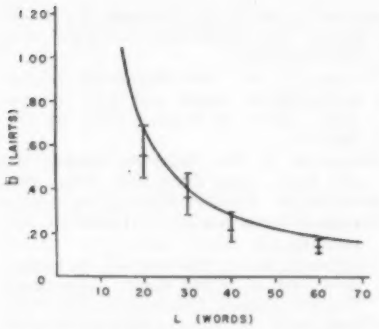

FIG 4. Predicted and obtained results for b a function of L with t constant.

In conclusion, it should be understood that, although not explicitly stated, all proposed relationships only hold within certain limits. Thus, it seems reasonable to assume that a list can be so long that S could not master it completely; then c would no longer be identical with L. Or, PT could be so long or so short that R_1 would no longer be a linear function of t. No serious attempt to determine limits has been made; however, the results just presented suggest that the fastest presentation times used may have been at or near the limits.

Also, it is obvious that when $t = 0$ then $R_1 = 0$, so literally it makes no sense to assume a positive value for the intercept m. However, for the sake of simplicity the intercept has been treated as if the function were linear throughout the range. In all probability m is the memory span; this is its rational interpretation and the numerical value agrees with the conventional values reported in the literature. The parameter k is a measure of the rate at which R_1 changes as a function of t.

Finally, we have no evidence as to why R is an exponential function of n, or why R_1 is a linear function of t. An understanding of why these relationships take the form they do awaits further research.

SUMMARY

A number of experiments are reported which investigated the immediate retention of unrelated words. Some studied retention following a single presentation; others studied the changes in retention that occurred with repeated presentations. The latter experiments were in effect studies of free-recall verbal learning. The method of free-recall verbal learning was selected for investigation because, it was found, the results of individual *Ss* learning single lists could adequately be described by an exponential function.

For the experiments on free-recall verbal learning a standard procedure and a standard method of fitting the exponential were used; normative data and data on reliability were presented. It was shown that there was no learning-how-to-learn or warm-up effect. Also, there was no difference between visual and auditory presentation or between individual and group testing. Learning was found to be a linear function of log frequency of usage.

The two experimental variables most intensively investigated were length of list and presentation time per item. The number of items retained after one presentation was a linear function of the total time required for presentation of the list, where the total time was a product of the length of list and the presentation time per item. Given this linear relationship and the exponential function for free-recall verbal learning it was possible to predict the learning of a list with a fair degree of accuracy given its length and presentation time.

REFERENCES

BENNETT, C. A., & FRANKLIN, N. L. *Statistical analysis in chemistry and the chemical industry.* New York: Wiley, 1954.

BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *J. gen. Psychol.*, 1953, **49**, 229–240.

BOUSFIELD, W. A., & SEDGEWICK, C. H. W. An analysis of sequences of restricted associative responses. *J. gen. Psychol.*, 1944, **30**, 149–165.

BOUSFIELD, W. A., SEDGEWICK, C. H. W., & COHEN, B. H. Certain temporal characteristics of the recall of verbal associates. *Amer. J. Psychol.*, 1954, **67**, 111–118.

BRUNER, J. S., MILLER, G. A., & ZIMMERMAN, C. Discriminative skill and discriminative matching in perceptual recognition. *J. exp. Psychol.*, 1955, **49**, 187–192.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.

DEESE, J. Serial organization in the recall of disconnected items. *Psychol. Rep.*, 1957, **3**, 577–582.

DEESE, J., & KAUFMAN, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. *J. exp. Psychol.*, 1957, **54**, 180–187.

EDWARDS, A. L. *Statistical methods for the behavioral sciences*. New York: Rinehart, 1954.

GREENBERG, R., & UNDERWOOD, B. J. Retention as a function of stage of practice. *J. exp. Psychol.*, 1950, **40**, 452–457.

HALL, J. F. Learning as a function of word-frequency. *Amer. J. Psychol.*, 1954, **67**, 138–140.

HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word probability. *J. exp. Psychol.*, 1951, **41**, 401–410.

MURDOCK, B. B., JR., & COOK, C. D. On fitting the exponential. *Psychol. Rep.*, 1960, **6**, 63–69.

PETERS, H. N. The relationship between familiarity of words and their memory value. *Amer. J. Psychol.*, 1936, **48**, 572–584.

SNEDECOR, G. W. *Statistical methods* (5th ed.) Ames: Iowa State Coll. Press, 1956.

STAGNER, R. Factors influencing the memory value of words in a series. *J. exp. Psychol.*, 1933, **16**, 129–137.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia Univer., 1944.

UNDERWOOD, B. J. *Psychological research*. New York: Appleton-Century-Crofts, 1957.

WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.

(Received September 21, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 4, 235–241

ANISOTROPY AND INTERACTION OF FIELDS OF SPATIAL INDUCTION

RICHARD M. MICHAELS[1]

Naval Research Laboratory

In a previous report (Michaels, 1957) spatial induction was measured about two yellow dots. There was a spread of activity in the retina about the projected image of the dots, and the magnitude of this excitation decreased exponentially about the borders of the figure. It was further shown that the fields about each of the figures interacted such that a unified field was developed. Thus, about the figures, there existed a strong field associated with and in close proximity to each dot, and also a secondary field which united the two figures into a coherent pattern. It was suggested that this secondary field represented a retinal correlate of such gestalt processes as continuity and grouping.

The present research was an attempt to examine the nature of the interaction of the fields more closely. It is of particular importance to establish some quantitative relation between the contribution of each dot to the induction at some particular location in the field. If the interaction effect is a process of linear addition, then it should be possible to predict the magnitude of the induction at some point between the figures by knowing the strength of the field of each dot as a function of distance. The present experiment was an attempt to test the validity of such a linear hypothesis.

A second problem deals with anisotropy in the induced field. According to Motokawa (1950) the magnitude of induction is greater in the horizontal dimension than in the vertical, and is thus anisotropic. In the previous study by Michaels (1957), anisotropy was not demonstrated for points close to the border of the figure. In order to test for anisotropy more carefully, the present experiments were done in both dimensions.

[1] Now with Bureau of Public Roads, Washington, D. C.

Apparatus and Procedure

The mechanics of this experiment were the same as presented previously (1957). Essentially, the inducing figures were one or two yellow dots 7 mm. in diameter each having a brightness of 2.6 ft.-L. These were presented for a period of 4 sec. and were followed after a 2-sec. interval by a small (1×1.2 mm.) rectangular white light. This was followed 1.5 sec. later by the electrical stimulus. Thus, this procedure measures the change in electrical threshold to the white light due to prior stimulation with yellow. By moving the white light to various points about the yellow figures the field of induction could be mapped.

In the first of these experiments the induction vs. distance function was determined about a single yellow dot. Upon the assumption of a symmetry in either the horizontal or vertical dimension, only three such curves were determined. One set of measurements was made along the vertical diameter, while the second was made along the horizontal diameter. The third measurement was along the 45° line intermediate between these two. From these curves the field of induction could be constructed about the yellow dot.

After the induction curves were obtained for the one dot, a second dot was added. The magnitude of induction was then determined for the single point along the horizontal diameter that lay intermediate between the two figures. After this point was determined, the procedure was repeated for two other dots having a greater separation between them. Four such pairs were used, and the distance between the two figures was 3.2,

9.2, 15.2, and 22.0 mm. Each point was replicated three times.

After the induction for the points in the horizontal dimension was determined, the pattern was rotated 90°. The measurements were then repeated in the vertical dimension. However, physical limitations required using an interdot distance of 16.2 mm. in place of the 15.2-mm. distance employed in horizontal measurements.

Thresholds were determined by a method of successive approximations as previously described (Michaels, 1957). The only change was that instead of varying the stimulus intensity in linear steps of current, logarithmic steps were used. The reasons for this change were twofold. First, the results of Riggs (1957) indicated that the threshold function was a normal distribution as a function of the logarithm of the current. Secondly, a log normal distribution is consistent with the psychological threshold relation for most intensity measures, e. g., sound intensity or brightness. From these considerations, induction was defined as $I = K(i_w - i_w')$ where i_w, in log μamps, is the electrical threshold to the white probe light when it is *not* preceded by the inducing figures; and i_w' is the electrical threshold to the white probe light when it is preceded by the inducing figures, again in log μamps. The constant K was arbitrary. This relation holds only for the case where the electrical threshold in the dark is constant for the two cases.

Two women were used as *Ss*, and both had normal vision. Each was tested individually two days a week for 4 hr. a day. In this period, 5–10 separate threshold determinations were possible. Two months of intensive training were required before stable thresholds were obtained. The sequence followed in the experiment was first to determine the electrical threshold in the dark. Next, the electrical threshold to the white light was determined, and finally the white light threshold was obtained when the yellow inducing figure preceded the white probe light.

Results

As was mentioned previously, the relation defining induction holds only so long as the dark threshold was constant. In order to validate this assumption, the dark threshold was determined once each day, either at the beginning of the day or the end.

TABLE 1

The Reliability of the Dark and White Light Thresholds

(K Log Current)

S	Dark Threshold		White Light Threshold	
	Mean	95% Conf. Int.	Mean	95% Conf. Int.
F. K.	2.07	±.04	1.88	±.04
C. T.	2.18	±.15	2.02	±.09

The distribution of dark threshold for each *S* is shown in Table 1. It seems reasonable to conclude that this threshold did in fact remain stable, and therefore the assumption in the induction equation was justified. Further, the threshold to the white light alone was also quite stable. The data, obtained over four months, are also shown in Table 1.

Measurement of the field of induction about one dot was obtained along the horizontal and vertical diameters, and the line intermediate. The intermediate curve was determined primarily to demonstrate that a continuous change in induction occurred between the two limits. The curves of induction vs. distance for the three cases are shown in Fig. 1. The curves do indicate that there is a greater

TABLE 2

Analysis of Variance of the Measurements around One Dot

Source	*df*	*MS*	*F*
Distance	4	5280.0	371.8*
Orientation	1	5568.0	392.1*
Ss	1	336.0	23.7*
D × O	4	57.0	4.0*
D × *Ss*	4	13.7	—
O × *Ss*	1	6.8	—
D × O × *Ss*	4	59.6	4.2*
Within	40	14.2	—
Total	59	—	—

* Significant at the .01 level.

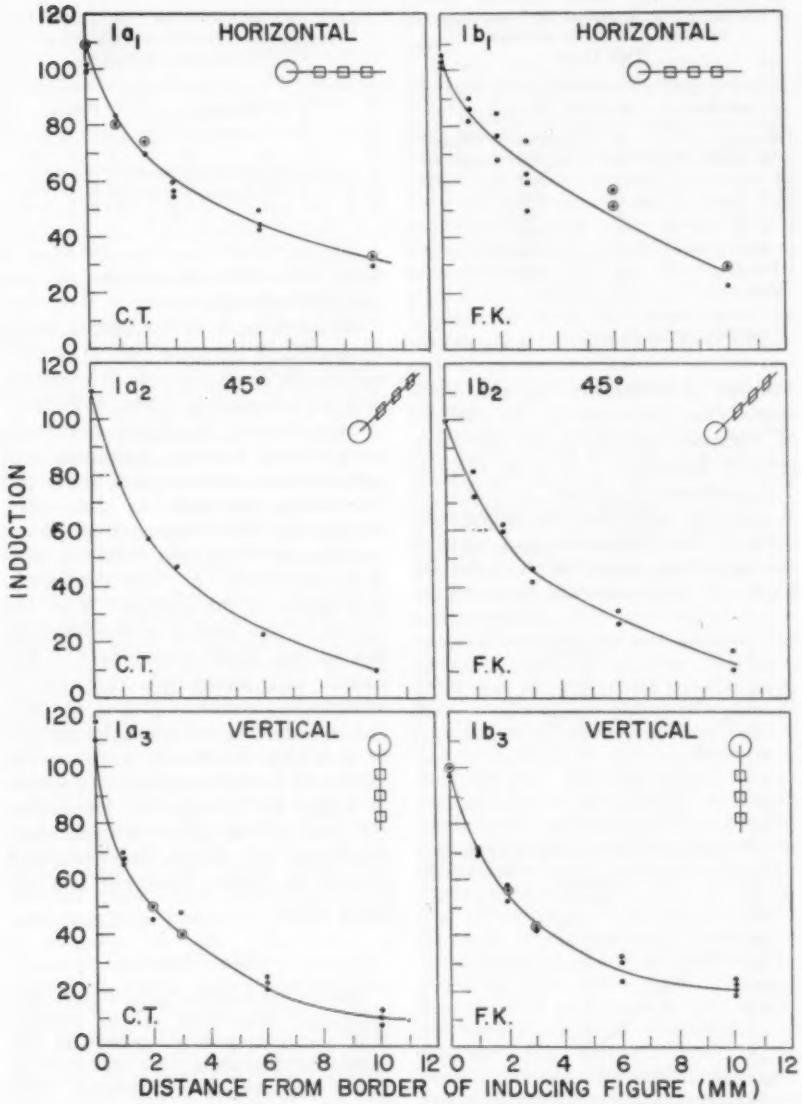

FIG. 1. Induction measured about one inducing figure.

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE OF DATA OBTAINED BETWEEN TWO DOTS

Source	*df*	*MS*	*F*
Distance	3	20,056	4964.0*
Orientation	1	17,623	4362.0*
Ss	1	768	19.0*
D × O	3	288.2	7.1
D × *Ss*	3	88.8	2.2
O × *Ss*	1	290.5	7.2
D × O × *Ss*	3	90.8	2.3
Within	32	40.4	—
Total	47	—	—

* Significant at the .01 level.

amount of induction for the horizontal than the vertical. An analysis of variance was done on this data for five distances, 1, 2, 3, 6, 10 mm. The summary is shown in Table 2. It may be seen that the differences between the horizontal and vertical are significant at better than the .01 level. It is reasonable to conclude

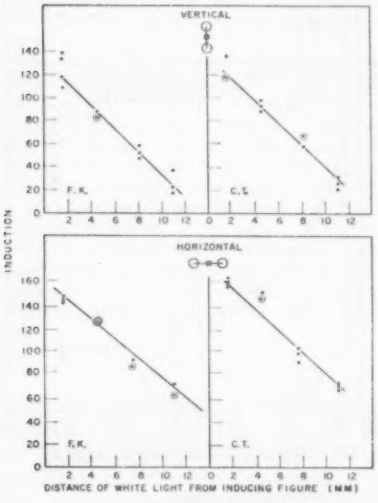

FIG. 2. Induction between two dots as a function of dot separation.

TABLE 4

SLOPE AND INTERCEPT OF EQUATIONS DERIVED FROM CURVES OF INDUCTION VS. INTERDOT DISTANCE

S	Horizontal		Vertical	
	a	*b*	*a*	*b*
C. T.	175.7	−9.3	138.5	−10.1
F. K.	165.3	−9.5	133.8	−9.9

from this that anisotropy of the induced field does occur.

Measurements of the field strength intermediate between two dots were submitted to an analysis of variance, and the summary is shown in Table 3. As may be seen, significant differences were found between distances and also between vertical and horizontal orientation as well as *Ss*. The interaction field measured at points equidistant from two inducing dots is shown in Fig. 2. The abscissa in this figure is the distance from the border of the border of the inducing dot to the white probe light. The points fall reasonably close to a straight line and the equation for each case was determined using the method of averages (Guilford, 1936). The constants in the equations are shown in Table 4. It may be noted that the two curves differ in both their intercepts and slopes, the horizontal having the higher intercept and the lower slope.

DISCUSSION

The results indicate quite clearly that anisotropy does exist in the induced field and induction is greater in the horizontal dimension than the vertical. In the previous study (Michaels, 1957) the field about the dots was presumed to be symmetrical and the iso-induction contours were drawn accordingly. It is necessary in light of the present data

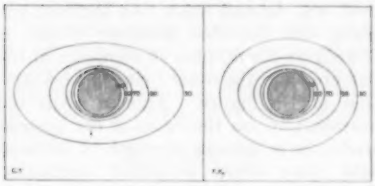

FIG. 3. Iso-induction contours about a circular pattern.

to revise this picture to take anisotropy into account. In Fig. 3 are drawn the iso-induction in contour about the inducing figure for each *S*. These figures were constructed assuming that each contour was an ellipse whose major axis was determined from the induction curve of Fig. 1. Similarly the minor axis was determined from the induction vs. vertical distance curves of Fig. 1. The ellipse as a description seems appropriate since the curve along the 45° line indicates a monotonic change in induction within the quadrant.

In Motokawa's (1950) study, it was stated that anisotropy was represented by a major to minor axis ratio of 1.2. The implication of this is that the two induction curves are identical in all respects save that the horizontal curve is displaced by a constant factor. That is, if the induction curve may be represented by the equation $I = Ke^{-x}$, then the only difference between the horizontal and vertical dimension is that the "K" is greater for one than the other. This conception seems highly unlikely on the basis of the present data. In the first place, both curves have the same maximum induction at the point of tangency ($d = 0$). Therefore, in order for anisotropy to appear, the slope of the two curves must differ and this appears to be the case in the present data as shown in Fig. 1. Furthermore, in the limit, the two curves should again converge as induction approaches zero. This again indicates that in the two dimensions, the induction equations differ in slope as well as in same constant factor. The magnitude of anisotropy cannot, therefore, be specified quantitatively without some specification of location on the induction vs. distance curve.

It was originally hypothesized that the field strength between two inducing dots should be linearly related to the field about one figure. Thus, for the two equal-intensity figures the field at a point intermediate between them should be twice the field induced by one. On the basis of this linear hypothesis a curvalinear relation should exist between induction and interdot distance. In Fig. 2 the relation, on the contrary, appears to be quite linear, and the magnitude of the induction is not simply related to the induced field about one dot. Thus, the hypothesis was not verified.

The reasons for the lack of correspondence between the data and a linear hypothesis are not clear. Actually, several reasons present themselves. First, threshold variability was considerably higher in these measurements than in the single dot case. This increased variability extends not only to repeat threshold measurements, but also the relative range of current required in order to meet the threshold criterion (Michaels, 1957) was greater in the two-dot case. Also, head and eye movements may be expected to influence the magnitude of induction more in the two-figure case than the one. Although a chin support was used, no further physical restrictions were placed on head position. Finally, and perhaps fundamentally, the measure of induction used was quite arbitrary. There is no reason, a priori, to assume that the induction is linearly related to the ratio of the electrical threshold to white light before and after stimulation of the eye with yellow light. Although this choice is an expedient one, it does not have much other justification. If this particular choice of scale is not linearly related to the underlying physiological phenomenon, then the hypothesis is not really being tested.

There are two features of the curves shown in Fig. 2 which are consistent

with the data from the single dot case. First, there is a constant term which displaces the induction in horizontal dimension by a factor of 1.2 from the vertical. This is consistent with the effects of anisotropy in the single dot case. The second point lies in the differences in slopes of the two curves. The decrease in field strength in the horizontal dimension is less as the two dots are separated than that occurring in the vertical dimension. Viewing the induction curves about the single dot, this result is what is to be expected, for the slope of the horizontal curve for one dot is less than that of the vertical curve. Thus, there is a consistency in the functions of the induction fields for the two cases. It seems reasonable to conclude that the induced field between two dots is directly related to the induction about a single dot. There does seem to be evidence for addition of fields, but the addition does not appear to be linear.

The results of the two dot experiment have implications for some previously reported pattern recognition studies (Shetler, Berbert, & Finney, 1955). In those studies, a pattern of colinear dots were masked by a background of random dots. It was found that the number of pattern dots required for detection was greater when the pattern to be detected was oriented vertically than when it was oriented horizontally. Further, detection of figures oriented at 45° lay between that of the other two orientations. The nature of the induced field suggests an explanation of these findings. The fact that a linear array of discrete dots constitute a distinct pattern indicates some kind of unification of these elements. The interaction of the induced fields between the dots represents one way in which such a gestalt comes into being, for the interaction field establishes a unifying link between the discrete elements. If this were true, then the "goodness" of such a pattern, and its resistance to destruction, should depend upon the magnitude of the interaction field. The magnitude of this induced field will depend upon two factors; one being the interdot distance which is related to the number of elements in the pattern; and the other being the orientation of the figure. Figures projected horizontally, on the basis of the present study, have greater strength of induced field than those projected vertically. Thus, a greater interdot distance (fewer dots) will be required in the horizontal dimension than in the vertical for detection to meet the same criterion. Since the strength of the induced field decreases continuously from a maximum in the horizontal dimension to a minimum in the vertical, those patterns oriented at 45° should have a field strength intermediate between the two extremes. Detection of the patterns so oriented should also be intermediate, and they appear to be so. The results of these pattern recognition studies are consistent with a model of grouping based upon spatial induction. It should be pointed out, however, that this correspondence is qualitative but not quantitative.

Summary

Using the electrical phosphene threshold as a measure of spatial induction, the magnitude of induction was determined about one yellow dot and between two. Anisotropy was determined in both cases by making measurements in both horizontal and vertical dimension. An additional induction curve was determined along the 45° line for the single figure so that a complete field could be constructed.

Results indicate that anisotropy does exist. The induced field was significantly stronger in the horizontal dimension than the vertical. The results were consistent for both the one- and two-dot experiments.

Measurements made between two inducing figures showed an apparently linear decrease in field strength as separation of the dots increased. On the basis of an hypothesis of linear addition of the fields, the magnitude of the induction should have been twice that of one dot and the relative with distance should have been curvilinear. The results do not support the hypothesis. Rather, they indicate simply that the interaction field is nonlinearly related to the induction about one dot.

The results on the interaction field about two dots were related to some previous studies on detection of patterns composed of discrete elements. The data on induction predicts the results of these pattern recognition studies.

REFERENCES

GUILFORD, J. P., *Psychometric methods*. New York: McGraw-Hill, 1936.

MICHAELS, R. M. The electrical phosphene threshold as a measure of retinal induction and visual organization. *J. exp. Psychol.*, 1957, **54**, 21–27.

MOTOKAWA, K. Field of retinal induction and optical illusion. *J. Neurophysiol.*, 1950, **13**, 414–426.

RIGGS, L. A., CORNSWEET, S. C., & LEWIS, W. G. Effects of light on electrical stimulation of the human eye. *Psychol. Monogr.*, 1957, **71**, (5, Whole No. 434).

SHETLER, W. L., BERBERT, J. H., & FINNEY, W. J. Some factors affecting the recognition of a linear pattern viewed against a random field. *J. Opt. Soc. Amer.*, 1955, **45**, 903. (Abstract)

(Received September 24, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 4, 242–244

REACTION TIME AS A FUNCTION OF STIMULUS INFORMATION AND AGE [1]

GEORGE J. SUCI [2]

Cornell University

MELVIN D. DAVIDOFF AND WALTER W. SURWILLO

National Heart Institute and Baltimore City Hospitals

With young *Ss*, reaction time is linearly related to bits of stimulus information (Hick, 1952; Hyman, 1953). Crossman and Szafran (1956) found nonlinearity [3] with old *Ss* when average response speed in a card-sorting task was used instead of reaction time. The present study tests the hypothesis that the regression of reaction time on stimulus information measured in bits is linear for a young and an old group of *Ss*.

For a variety of processes, as difficulty increases, age decrement also increases (Welford, 1958). Taking increased stimulus information as a measure of increased difficulty, a second hypothesis is that age decrement in reaction time will increase as a function of greater amounts of information.

METHOD

Subjects.—Two age groups were used; each consisted of 12 males. The median age of the older group was 63.0 yr., range 60–70; the median age for the younger group was 18.5 yr., range 17–38. Of the older *Ss*, five held a doctorate, five had a master's degree, and two had a bachelor's degree. In the younger group one *S* had a master's degree and the rest were currently college undergraduates. No *S* had any known major or relevant (to this task) physical or psychological handicap. The older *Ss* were noninstitutionalized people living in the community.

Apparatus.—Sixteen lights were mounted as a square matrix, $4\frac{1}{2} \times 4\frac{1}{2}$ in., on flat, black plywood tilted 25° from the vertical away from *S*'s position. Only the four corner lights were used. The lights, 120 v., 6 w., were shone through $\frac{13}{16}$-in.-diameter green jewel caps. A .01 sec. Standard Electric timer was automatically started when *E* turned off any one of the lights and automatically stopped by *S*'s verbal response. All other timing was done with a .2 sec. stop watch. The *E* was screened from *S* with a gray cardboard panel. The *S* sat approximately $3\frac{1}{2}$ ft. in front of the lights.

Procedure.—The *Ss* learned to designate each of the four lights with BEP, BIX, BUJ, BOZ, respectively. After preliminary training, *S* was instructed to get set to respond when *E* said "ready." At the ready signal all four lights went on simultaneously; 2 sec. later one of the lights went off. The *S* was instructed to name the light that went off as rapidly as he could. The *S* was given the correct response if he erred or if he failed to respond after 5 sec.; he was told "right" after making a correct response.

The lights were presented in random order with the restriction that each light was the stimulus 3 times in every 12 trials. The trials were presented in blocks of 24 followed by a 3-min. controlled rest period. During the rest period *Ss* were given a word association test. The learning session ended when *S* responded correctly to an entire block of trials.

The second part of the experiment consisted of three blocks of 24 trials each, representing the 3-, 2- and 1-light conditions. Each block of trials was followed by a 5-min. uncontrolled rest period consisting of conversation with *S*. The order of presentation

[1] Grateful acknowledgment is made to Sally Stram and Jesse Yaffa for their able assistance in the laboratory.

[2] This research was performed while this author was with National Institutes of Health, Gerontology Branch, Baltimore City Hospitals, Baltimore, Maryland.

[3] Crossman and Szafran (1956) furnish no statistical evaluation of the linearity of this function.

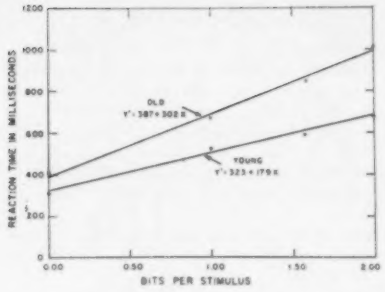

FIG. 1. Reaction time as a function of stimulus information and age.

of the conditions varied through the six possible permutations twice. When a very occasional error was made in this unit of the procedure, the stimulus involved was represented after going on with a small number of the other stimuli.

The procedure with three lights will illustrate the procedure used with all conditions. Three of the four lights appeared "on" when the ready signal was given. There are four ways of selecting three out of four lights. Each combination appeared an equal number of times in the block of 24 trials. Two, three, or four seconds after the ready signal, one of the three lights went off to be named by S. The preparation time was varied to minimize anticipatory responses, especially in the 1-light condition. Each of the four stimulus lights and three preparation times was used an equal number of times in the block. The presentation of combinations, stimuli, and preparation times were randomly distributed throughout the 24 trials.

RESULTS

The mean of the median values for each *S*'s reaction time to each stimulus information condition was plotted against the bits of stimulus information in Fig. 1. The 1-, 2-, and 3-light conditions represent 0.00, 1.00, and 1.58 bits of information, respectively; the final block of trials in the learning session was used to obtain reaction times in a 4-light or 2-bit condition. Each median is based on 24 errorless trials. The number of errors was no greater than two for any *S* in the total of 24 trials.

As the small F values indicate (Table 1), the hypothesis of linearity of regression could not be rejected for either age group. The slopes of the regression were .302 and .179 for old and young, respectively. The difference between the obtained slopes was evaluated by the t test (Edwards, 1954, pp. 310–312) and found to be significant with $t = 4.40$, $df = 92$, $P < .01$.

The assumption of homogeneity of variance and normal distribution could not be tested and an alternative statistical test was applied. A regression line was fitted by the least-squares method to each *S*'s distribution of reaction times on stimulus information. The slopes of these regressions were compared for young and old groups with the Mann-Whitney U test (Siegel, 1956, pp. 116–127). The null hypothesis was rejected ($P < .001$) and the significant differences between slopes found by the t test was supported.

A surprising result was that the older *S*s reached the learning criterion in fewer trials than the younger *S*s. Mowbray and Rhoades (1959) indicate that increased amounts of practice tend to increase the slope of reaction time on information. By comparing

TABLE 1

ANALYSIS OF VARIANCE FOR LINEARITY OF REGRESSION

Age Group	Source	*df*	*MS*	*F*
Old	Deviation of means from regression line	2	.009	.529
	within arrays	44	.017	
Young	Deviation of means from regression line	2	.004	.800
	within arrays	44	.005	

the six slowest learners of the older group to the six fastest learners of the younger group, the age difference remained constant but the amount of practice was reversed. The same statistical tests were applied and the results were unchanged. Thus, practice effect differences were ruled out as explanatory of slope differences.

Summary

Two hypotheses were tested: (*a*) reaction time is a linear function of stimulus information both in old and in young *Ss*; and (*b*) age differences in reaction time increase as a function of increasing amounts of stimulus information. Reaction times to a stimulus of one light-off in subsets of one light (0.00 bits of information), two lights (1.00 bit), three lights (1.58 bits) and four lights (2.00 bits) were measured. Both hypotheses were supported by the results.

References

Crossman, E. R. F. W., & Szafran, J. Changes with age in the speed of information-intake and discrimination. In F. Verzar (Ed.), *Experimental research on ageing*. Basel: Birkhäuser, 1956. Pp. 128–135.

Edwards, A. L. *Statistical methods for the behavioral sciences*. New York: Rinehart, 1954.

Hick, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4**, 11–26.

Hyman, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.

Mowbray, G. H., & Rhoades, M. V. On the reduction of choice reaction times with practice. *Quart. J. exp. Psychol.*, 1959, **11**, 16–23.

Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

Welford, A. T. *Ageing and human skill*. London: Oxford Univer. Press, 1958.

(Received September 25, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 4, 245–249

RETROACTIVE INHIBITION OF CONNECTED DISCOURSE AS A FUNCTION OF SIMILARITY OF TOPIC [1]

NORMAN J. SLAMECKA

University of Vermont

The few attempts that have been made to study the effect of degree of similarity between original learning (OL) and interpolated learning (IL) upon the retroactive inhibition (RI) of prose passages have utilized tests of substance or of factual retention (Ausubel, Robbins, & Blake, 1957; Hall, 1955; McGeoch & McKinney, 1934). All of these studies have reported either negative (no significant RI) or inconclusive findings. We have been unable to find any published reports examining the effect of this variable on the rote retention of such materials, in spite of the almost exclusive use of rote retention measures with nonsense materials.

Recent preliminary research from this laboratory (Slamecka, 1959) has indicated that the rote retention of a passage of connected discourse decreased directly as the topic of an interpolated passage is closer to the topic of the original passage. That is, as the subject matter of two succesively learned passages was made more similar, the delayed recall of the first passage became progressively poorer. Such a finding is in accordance with results that have been obtained by the use of unconnected materials (McGeoch & McDonald, 1931; Osgood, 1946), and implies that the retention of a prose passage can be predicted in much the same manner as that of the classical serial list, at least as far as the variable of similarity is concerned. Before such a conclusion is accepted, however, certain limiting aspects of the previous experiment should be pointed out; namely, only one original passage was presented, *S*s were tested in groups, the methods of whole presentation and written recall were used, passages were presented for a fixed number of trials, and the data were based on a unique scoring system. Recognizing the potential practical importance of a soundly established knowledge of the determiners of the RI of prose, it was felt that a yet more stringent test of the similarity hypothesis was indicated, in order to verify and extend the generality of the findings to other materials and methods. The present study provided such a test, in that; (*a*) several original passages were presented, (*b*) *S*s were tested individually, (*c*) the serial anticipation method was used, (*d*) acquisition was to a performance criterion, and (*e*) error analyses were performed. The hypothesis tested was that the RI of the verbatim recall of prose sentences is a positive function of the degree of topical similarity the interpolated passage bears to the original passage.

METHOD

Subjects.—The *S*s were 28 undergraduate students who participated as part of the requirements of an introductory psychology course. Each one took part in five sessions, the first of which was a practice period designed to familiarize *S* with the general procedure and to reduce practice effects. Every *S* was tested under all the four experimental conditions.

[1] This research was supported by NSF Grant G-6192. The author wishes to express his gratitude to Ernest Damianopoulos, who collected the data.

Materials.—Four original passages were chosen from a single book source and modified slightly so that each consisted of a single sentence 20 words in length. Six other 20-word sentences were obtained for potential use as the IL passages. Six faculty members and graduate students served as judges in making independent ordinal rankings of these six passages into the categories of High, Intermediate, and Low similarity of topic to each of the four original passages, by putting two passages under each category. The judges' rankings were unanimous. Next, in order to determine the difficulty of these passages, all six were learned to a criterion of one perfect anticipation by each of 12 *S*s. They were presented in a counterbalanced order over three sessions, using a 3-sec. presentation rate and a 6-sec. intertrial interval. Three passages, one from each similarity category, were chosen as being equivalent in difficulty. The mean trials to criterion for these passages were: High—7.9, Intermediate—7.9, and Low—8.7. The obtained *F* for difficulty was less than one. As a further check, these three passages alone were learned to the same criterion by each of another group of 12 *S*s. Each *S* learned all three in one session in a counterbalanced order, with the same rates of presentation as before. Again, *F* for difficulty was less than one. The respective mean trials to criterion were: High—8.2, Intermediate—7.2, and Low—7.4. The passages are reproduced below.

Original 1: So habitual become our expectations about symbols invariably possessing referents, that we tend unconsciously toward such assumptions concerning every word. *Original 2:* However in most instances, linguists are predominantly concerned about determining the concrete denotations of discrete words functioning as proper names. *Original 3:* We must postulate that from strictly semantic points of vantage, most confusions in communication revolve about inadequate stipulation of meaning. *Original 4:* Much evidence regarding the inference of authentic meanings which are not stipulated, is derived from context analysis within the paragraph.

Interpolated, High Similarity: Communicators can exercise latitude in specifying meanings however they choose, provided that such definitions correspond somewhat closely to customary usage. *Interpolated, Intermediate Similarity:* These results appear to be inconsistent with the finding that mathematics has proved itself eminently applicable to empirical scientific knowledge.

Interpolated, Low Similarity: This government eventually achieved independence against nearly overwhelming resistance, and maintains national integrity only at the price of unremitting vigilance.

It is obvious that the topic of all the original passages, as well as the Interpolated High Similiarity passage, was *semantics*, that of the Interpolated Intermediate Similarity passage was *mathematics*, and that of the Interpolated Low Similarity passage was *government*.

Procedure.—Following the initial practice session each *S* took part in four experimental sessions, at about two-day intervals, which were conducted according to the usual RI paradigm. Each session began with two trials of color-guessing warm-up activity, the colors appearing on the memory drum. Appropriate instructions for serial anticipation were given, and an OL passage was then presented. After OL had been carried to one perfect trial, an IL passage was given and also learned to one perfect trial. In all cases the words were presented at a 3-sec. rate with a 6-sec. intertrial interval. Since both OL and IL were to a performance criterion, all the passages were equivalent in degree of mastery. The time between end of OL and the start of relearning (RL) of the original passage was held constant at 12 min., 6 sec., which was equivalent to 12 trials at the rates used. This period was fixed on the basis of preliminary work, and proved sufficiently long to allow all *S*s to achieve criterion on the IL. For the control condition (no IL) and for *S*s who took less than 12 IL trials, a filler task of color guessing was given in order to prevent rehearsal while still maintaining the set of responding at the memory drum. OL retention was measured in four RL trials, and was followed by four RL trials on the interpolated passage, the latter being done to avoid biasing *S*s toward one of the lists.

The four sessions differed only in the condition of the interpolated activity; High, Intermediate, Low, or Control. The sequence of conditions was counterbalanced among *S*s, and also each original passage was paired with every interpolated condition an equal number of times, producing, with 28 *S*s, a total of seven replications.

Results

Table 1 presents the results of the OL and IL acquisition. Although there is as yet no standard procedure for testing the homogeneity

TABLE 1

MEAN TRIALS FOR ORIGINAL AND INTERPOLATED LEARNING

Measure	Similarity			
	High	Int.	Low	Control
OL Mean	9.8	10.0	10.7	11.1
SD	3.6	3.6	3.6	3.9
IL Mean	7.6	6.3	7.0	
SD	2.7	2.4	2.1	

of variance of correlated samples, an inspection of the *SD*s of Table 1 indicated very little difference in variability among the groups. Briefly, there were no significant differences in the trials taken to learn either the original or the interpolated passages. A slight amount of associative facilitation may be present in the IL condition, since each passage was learned in fewer trials during the experiment than during the difficulty tests. However, this transfer does not appear to be a function of similarity level.

The data for recall, the performance on the first RL trial, revealed a highly significant effect of similarity upon retention in the direction predicted ($F = 17.6$, $df = 3$, $P < .01$), and Tukey's test gave a significance gap of 1.5 (.05 level, two-tail). Friedman's test gave a χ_r^2 of 28.23,

TABLE 2

MEAN RECALL SCORES FOR RELEARNING TRIALS

Cond.	Relearning Trials							
	1		2		3		4	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
High	11.9	4.1	17.0	2.5	18.7	1.6	19.3	1.3
Int.	13.5	2.9	17.9	1.9	19.2	0.9	19.0	1.3
Low	15.0	2.2	18.4	1.4	18.9	1.5	19.2	1.2
Cont.	17.2	2.6	19.2	1.1	19.4	1.2	19.7	0.5

TABLE 3

ANALYSIS OF ERRORS AT RECALL

Cond.	% Intrusions	% Omissions	Overt Responses	
			% Total	% Correct
High	7.9	32.7	67.3	88.3
Int.	8.6	24.1	75.9	88.7
Low	8.0	16.8	83.2	90.3
Cont.	6.1	7.8	92.1	93.4

which at 3 *df* is significant beyond the .001 level. Table 2 shows the entire relearning data for all conditions, indicating the very transitory nature of RI, which is in accord with findings based upon nonsense materials. The differences among the groups found at recall persisted through the second RL trial, but by the third RL trial the differential effects of similarity had been overcome.

Table 3 presents the tabulation of errors at recall. Since overt intrusions in connected discourse cannot be unequivocally assigned to various specific subcategories, all errors were subsumed under only two headings; omissions (failure to respond) and intrusions. The table also includes a count of all overt responses (whether correct or incorrect), and the percentage correct of the latter.

DISCUSSION

As all of the work groups differed significantly from each other as well as from the control's baseline, absolute RI of recall was clearly obtained. Relative RI values $\left(\frac{\text{Rest-Work} \times 100}{\text{Rest}}\right)$ for recall also showed the predicted directionality of effects as follows; High—30.8%, Intermediate—21.5%, and Low—12.8%. These findings verify and extend the generality of the results of the previous study to other passages and to the serial anticipation method.

The way in which such a nonlocalized variable as similarity of topic operated to produce these results, poses a problem. It is evident that this is not a specific interference effect as is the case with an A-B, A-C paired-associates design. Rather, we must postulate a more general, nonspecific kind of interference based upon previously acquired language habits. A consideration of the nature of the task facing *S*, together with an analysis of the errors at recall, may afford a tentative explanation of the processes underlying the observed decrements in retention. We hypothesize that in memorizing a passage of connected discourse, *S* proceeds by emitting responses from a progressively narrower range of possible responses. As learning progresses, *S*'s responses, instead of being emitted more or less at random, become increasingly restricted to those that are congruent with the conceptual class corresponding to the "topic" of the passage, since it is only words of that class that are being reinforced or correctly anticipated. The subjective counterpart of this process of response-range restriction through selective reinforcement is probably *S*'s feeling that he is beginning to grasp the "meaning" or the general "sense" of the passage, and he may then be able to paraphrase the passage by emitting words from the same conceptual pool. The original passages for this study dealt with the topic of semantics, and it is to that class of words comprising the semantics pool that *S*'s responses ultimately become narrowed. Some such process must be occurring since *S* does in fact learn the passage, and there is no reason to suppose that his functioning response repertoire was at all similarly restricted at the start of acquisition. This conceptual response restriction, together with the added contextual constraints arising from the grammatical structure of the language, reduces the probability of response errors, and may account in large part for the well-known rapidity with which connected discourse is learned, as compared to the difficulty in mastering a nonsense list of equal length.

Having learned two passages in succession, *S* must next attempt to recall the first one. A competition-of-response formulation with stress upon the discriminability of the responses appears most promising for incorporating the recall data, particularly for those shown in Table 3. Responses from both OL and IL are in conflict for emergence at recall and *S* must discriminate correctly between the competing alternatives on the basis of which one corresponds most closely to the proper conceptual class.[2] It has long been known that the latency of discriminative responses is a positive function of the degree of similarity between the stimuli to be discriminated (Woodworth & Schlosberg, 1954). In the present study, the greatest similarity between competing response alternatives should obtain under the High condition, where responses for both OL and IL are referrable to the same conceptual pool. Therefore, the response latencies should be longest for the High condition, and decrease consecutively for the Intermediate, Low, and Control conditions, whose items come from increasingly different response classes. Since *S* has only 3 sec. in which to give a response (the rate of presentation used in this study), it follows that errors of omission (reflecting latencies greater than 3 sec.) should be most probable at the High condition and decreasingly probable in the aforementioned order of the remaining conditions. That is the interpretation given to the data in the second and third columns of Table 3, where it is seen that as the IL similarity increased there was an increase in the percentage of overt responses attempted. The

[2] This view bears some resemblance to Underwood's (1945) "differentiation" construct, wherein the differentiation of list membership of competing responses is stressed. However, Underwood has not formally extended that construct to an analysis of the effects of similarity relationships, and further, he has used it to account only for the pattern of errors in RI, whereas the present formulation attempts to account not only for the error distribution but also for the amount of RI.

fourth column shows that regardless of the particular experimental condition, a relatively constant percentage (about 90%) of all responses attempted were correct. This is interpreted as a reflection of the fact that all the passages were carried to the same degree of mastery and therefore, accuracy of recall should be about equivalent (also see Slamecka, 1960). Thus, it is suggested that the effect of the similarity variable is not to produce a weakening or unlearning of the original associations (since the accuracy of those responses attempted remained constant over all conditions), but rather to lengthen their response latency, thereby increasing omission errors and producing an apparent retention decrement.

A straightforward test of the above hypothesis could be made by repeating the present experiment but allowing unlimited response time at recall. The hypothesis would predict that there would be no differential RI among the work groups, but that correct response latencies would be a positive function of IL similarity level.[3]

Summary

It was hypothesized that the retroactive inhibition of the rote retention of prose passages was a positive function of the degree of similarity of topic the interpolated passage bore to the interpolated passage. Three levels of judged topical similarity were employed, and significant RI was obtained. The hypothesis was confirmed, thus verifying and extending the generality of previous findings. A tentative formulation stressing response competition and increased latencies was discussed.

[3] The use of latency measures in RI work is rare, but encouraging results were obtained by Osgood (1946) who reported a relation between response times and degree of response similarity in an A-B, A-C design.

References

AUSUBEL, D., ROBBINS, L., & BLAKE, E., JR. Retroactive inhibition and facilitation in the learning of school materials. *J. educ. Psychol.*, 1957, **48**, 334–343.

HALL, J. F. Retroactive inhibition in meaningful material. *J. educ. Psychol.*, 1955, **46**, 47–52.

MCGEOCH, J. A., & MCDONALD, W. T. Meaningful relation and retroactive inhibition. *Amer. J. Psychol.*, 1931, **43**, 579–588.

MCGEOCH, J. A., & MCKINNEY, F. The susceptibility of prose to retroactive inhibition *Amer. J. Psychol.*, 1934, **46**, 429–436.

OSGOOD, C. E. Meaningful similarity and interference in learning. *J. exp. Psychol.*, 1946, **36**, 277–301.

SLAMECKA, N. J. Studies of retention of connected discourse. *Amer. J. Psychol.*, 1959, **72**, 409–416.

SLAMECKA, N. J. Retroactive inhibition of connected discourse as a function of practice level. *J. exp. Psychol.*, 1960, **59**, 104–108.

UNDERWOOD, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59**, (3, Whole No. 273).

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, 1954.

(Received October 2, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 4, 250–254

SHOCK STRENGTH, SHOCK REDUCTION, AND RUNNING SPEED[1]

JOHN P. SEWARD, RICHARD A. SHEA, ARTHUR A. UYEDA, AND DAVID C. RASKIN

University of California, Los Angeles

The basic problem was how drive and reward combine to determine response. In aversive behavior the problem can be stated in terms of intensity of a noxious stimulus and the degree to which the response reduces it. Specifically, how does the level of a drive-stimulus modify the reinforcing effect of reducing it by a fixed amount?

Campbell and Kraeling (1953) attacked this question by measuring the speed of rats in an electrified runway with different intensities of shock in the main- and end-sections. They found something analogous to the Weber-Fechner principle, in that the same amount of shock reduction was more effective at lower shock levels than at higher ones. An opposite relation, if any, seems to hold for appetites, where the effect of a constant food-reward is sometimes enhanced by increased hunger (Seward, Shea, & Davenport, 1960; Seward & Procter, 1960; Seward, Shea, & Elkind, 1958). In view of this discrepancy, Campbell and Kraeling's method was adopted in essentials to seek further data on the interaction of drive and reward in escape.

METHOD

Apparatus.—The runway was 4 ft. long by 3¾ in. wide; startbox and endbox were 10 × 5½ in. The floor was a grid of 1/16-in. metal rods ½ in. apart. The runway was roofed with hardware cloth and the two boxes with Plexiglas 2½ in. above the grid. A photocell was mounted 6 in. from each end of the runway. A Plexiglas guillotine door between startbox and runway closed a microswitch and activated two timers, one controlled by each cell. By closing another switch *E* could turn on a buzzer just under the startbox and a 7½-w. light just over it as a signal of impending shock. The experiment was run in a darkened room.

The shock source was the matched-impedance circuit described by Campbell and Teghtsoonian (1958) for the reasons they discussed. Sixty-cycle current from a step-up transformer was controlled by a potentiometer and fed to the grid through a 150,000-ohm fixed resistor. A microammeter in series with the grid gave readings of current. Two identical stimulators were used, one connected with the startbox and runway, the other with the endbox.

Subjects.—The *Ss* were 108 rats in 12 replications. All were Sprague-Dawley albinos, with few exceptions between 3 and 6 mo. old, and all but three replications were female. Food and water were always available in the living cages.

Design.—Four levels of shock were used in the runway—315, 255, 190, and 125 v.—and three in the endbox—190, 125, and 0 v. There were nine combinations, of which eight are shown in the left-hand column of Table 2; the ninth was 125-0. Higher voltages had been tried in preliminary work but proved too traumatic. The *Ss* had also been tested at several levels with the same voltage in runway and endbox, but all refused to run within a few trials. A 0-0 treatment was dropped for the same reason.

The nine *Ss* in a replication were assigned at random to the nine treatments. Each replication was handled by one *E*.

Procedure.—Two days were allowed for adapting to the apparatus. On Day 1 *Ss* were given 30 min. to explore it in groups of four or five. On Day 2 they had four timed trials in rotation without shock.

Either next day or the day after, *Ss* were tested in random order, four on one day, five on the other. A test consisted of 30 massed trials, the first and last of which did not count. Before starting, *E* set the

[1] Aided by a grant (G 4419) from the National Science Foundation.

potentiometers at the required voltages but made only the endbox "hot." On Trials 2 through 29 *E* put *S* in the startbox with the grid cold, pressed the buzzer-light switch, and about 1 sec. later turned on the shock and opened the door simultaneously. Thirty seconds after *S* entered the endbox *E* turned off all shock, removed *S*, and put it in the startbox for the next trial. Every fourth trial, if not more often, he read the endbox current. Trials 1 and 30 were used to read current in the startbox. After applying shock *E* kept the door closed until he had obtained a reading. If, as happened occasionally, *S* refused to enter the endbox within 60 sec., its feet were irritated from below with a stiff brush. Four such refusals in succession terminated the test.

Results and Discussion

Measures.—Drive strength was defined as the total voltage from the potentiometer; drive reduction, as the difference in voltage between runway and endbox. These definitions are arbitrary, since it is not known what characteristic of shock is most effective (Campbell & Teghtsoonian, 1958). Also unknown is the relation between stimulus intensity, however measured, and drive strength, though Campbell (1956) has made a start in that direction. Total voltage was chosen, first, because it was the independent variable actually manipulated; as such it was constant for all members of a group. Secondly, the selected voltages permitted comparisons between groups given the same reduction from different levels of shock. Amperage readings were obtained, however, making it possible to compute mean current, bodily resistance, and power across the body for each voltage used. Since others may find these measures useful they are presented in Table 1.

Two independent measures of performance were available: *starting time*, given by the first clock, and *running time*, given by the difference between the first clock and the second. Each

TABLE 1

Mean Resistance, Current, and Power Corresponding to Applied Voltages

Volts	Rat's Resistance	Micro-amperes	Milliwatts
315	64,000	1470	138
255	98,300	1030	104
190	113,000	720	59
125	136,700	440	26

S's 28 trials were divided into 7 blocks of 4; the median time of each block was converted into its reciprocal times 10. Transforming time into speed scores had the advantages of normalizing the distributions and minimizing the inaccuracy introduced by the few trials interrupted at 60 sec.

Data will be reported for only eight of the nine treatments compared. In the group run from 125 to 0 v. just 3 of the 12 *S*s continued to run for 30 trials; 7 failed to complete a single block of 4.[2]

Starting speed.—This measure will be dealt with summarily, since it yielded little of interest. Mean starting speeds appeared to vary inversely with endbox voltage and to rise and then fall as startbox voltage increased. There was enough consistency to suggest real effects; in particular, to judge by *S*s' behavior, 315 v. in the startbox had a disruptive effect that interfered with efficient escape. But the trends were not reliable. Analysis of variance, Treatments × Replications × Blocks of trials, showed Blocks to be the only significant effect ($P < .001$). Scores improved continually with

[2] Since 440 μa. are well above the rat's threshold, this result is believed not to be due to weak current alone. It was probably more painful to walk on the grid than to stand still; even the shock-free endbox was not worth the added discomfort of getting there (see Campbell & Teghtsoonian, 1958, p. 191).

TABLE 2

MEAN RUNNING SPEEDS (10 × RECIPROCAL OF TIME) IN BLOCKS OF FOUR TRIALS

Volts		Blocks			Mean
Runway	Endbox	1	4	7	
315	0	13.7	14.0	13.4	13.6
315	125	12.2	13.3	12.4	13.0
315	190	9.5	10.5	12.0	10.5
255	0	11.9	12.7	12.1	12.2
255	125	11.0	12.5	12.1	12.1
255	190	8.3	9.2	8.9	8.9
190	0	10.4	11.6	11.9	11.3
190	125	8.3	8.6	6.8	8.0
Mean		10.7	11.6	11.2	11.2

practice regardless of shock strength. All other effects were obscured by exceedingly high variance in the error terms.

Running speed.—Table 2 gives mean running speeds for the first, middle, and last blocks of trials; these are typical of the data as a whole. To evaluate the differences in rows and columns an analysis of variance was performed using all the data and is presented in Table 3.

It will be seen that the effects of Blocks and of their interaction with Treatments are both highly significant. As shown in Table 2, all eight groups increased speed from the first block to the fourth; six declined from the fourth to the seventh. Both trends are probably real (a difference of 0.4 between-block means is significant at the .05 level) and are presumably due to practice and habituation, respectively. Since the absolute fluctuations from block to block were small, however, and there was no a priori ground for selection, further analysis was based on the over-all means.

TABLE 3

ANALYSIS OF VARIANCE OF RUNNING SPEEDS

Source	*df*	*MS*	*F*
Between *Ss*			
Treatments (T)	7	320.6	7.94*
Replications (R)	11	128.1	
T × R	77	40.4	
Within *Ss*			
Blocks (B)	6	9.2	4.60*
B × T	42	4.8	3.20*
B × R	66	2.0	
B × T × R	462	1.5	

* $P = .001$.

Differences among treatment groups were also significant at the .001 level. The next step was to compute critical differences for the two-tailed t test applied to pairs of groups. For this purpose, since group variances were homogeneous, it was possible to use the mean square for interaction with replications, with 77 *df*, as an error term. It was found that to reach significance at the .05 and .01 levels two groups would have to differ by 2.0 and 2.6, respectively. These criteria, applied to the last column of Table 2, show that at the two higher drive levels speed was significantly reduced only by an endbox shock of 190 v.; at the lowest drive level 125 v. were enough.

Chief concern, however, was not with endbox voltage as such but with amount of drive reduction. The means of Table 2 were therefore plotted against differences between runway and endbox voltages, as shown in Fig. 1. Here interest centers on the vertical differences; i.e., between groups running at different drive levels to the same "reward." There are three such pairs, none of which quite reaches the .05 criterion of significance. (The differences of 1.6 and 1.7 at voltage differences of 125–130 and 190 are significant at about the .10 level.)

We are faced with a number of possible alternatives. First there is the choice between accepting and rejecting the

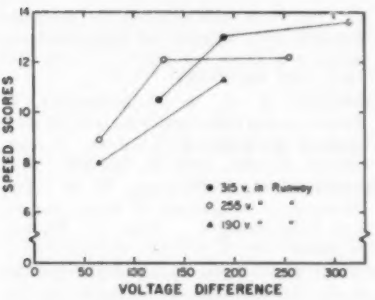

FIG. 1. Mean running speeds at each drive level plotted against voltage differences between runway and endbox.

null hypothesis, according to which the plotted points would represent deviations from a single curve relating speed to drive reduction. If the decision is to reject, a second question is whether the "true" curves are parallel or not.

Aside from Campbell and Kraeling's finding (1953), several factors make it difficult to accept the null hypothesis:

1. Curves drawn through the points for separate drive levels would reach different asymptotes. The reason is that the parameters of the points farthest to the right *are* the asymptotes, in the sense that they correspond to maximal drive reduction. Note that at those points the difference between 315 and 190 v. is significant.

2. Assuming that such curves would reach the baseline before the voltage difference was reduced to zero, we should not expect them to reach it at the same point. This prediction is based, not on extrapolation, but on Campbell's (1956) finding that the "reinforcement difference limen" for shock reduction increases with shock strength over the range of voltages here used. The curve for 190 v. should therefore meet the baseline closest to zero, 255 next, and 315 farthest away.

Points 1 and 2 can both be true, however, only if the "true" curves for different drive levels intersect somewhere in their courses; i.e., only if drive and drive reduction interact.

3. Although no obvious measure of interaction was available, the data provided a suitable makeshift. From Fig. 1 it is clear that interaction could be evaluated if a group had been tested with 255 v. in the runway and a drop of 190 v. in the endbox. Since the curve at that level becomes almost horizontal, the data for the last point could be substituted with little distortion. This was done, and a 2×2 analysis of variance was performed on four groups representing two levels of drive (255 and 315 v.) and two levels of drive reduction (125–130 and 190 v.). Tested against the over-all error term in Table 3 the interaction reached only the .10 level, but against its own error term (the second-order interaction with replications) it was significant at .05.

Of the three alternatives in question—a single curve, parallel curves, and nonparallel curves—these considerations point toward the last. If they carry weight, the present data do not conflict with Campbell and Kraeling's conclusion that the effect of drive reduction is a function of drive level. The nature of the function, however, remains uncertain. Campbell and Kraeling (1953) asserted that "a constant reduction in shock is more effective at a low initial shock level than at a high initial shock level" (p. 101). Our results suggest that as shock reduction increases a point is reached where their statement no longer applies and may even be reversed. It is of interest that if their asymptotic data are plotted as in Fig. 1 the curves converge and two of them actually meet. To permit a definite formulation of wide generality, more points must obviously be plotted with greater precision.

SUMMARY

To determine the relation between shock strength and shock reduction in motivating escape, 108 rats were tested for speed of running from one of four voltages (315, 255, 190, and 125) in a runway to one of three (190, 125, and 0) in an endbox.

Starting speeds showed a significant practice effect over 28 massed trials, but treatment differences were obscured by high variance.

Running speed showed a small but significant change over trials. With no charge in the endbox, mean speed dropped significantly when runway shock was reduced from 315 v. to 190. At higher levels it dropped significantly when endbox shock was raised to 190 v. A charge of 125 v. in the runway failed to maintain escape behavior; in the endbox it had a retarding effect only with 190 v. in the runway.

With shock reduction held constant, differences between initial shock levels were of doubtful significance. Other indications, however, pointed to a relation between effect of shock reduction and initial intensity, the precise form of which is still obscure.

References

CAMPBELL, B. A. The reinforcement difference limen (RDL) function for shock reduction. *J. exp. Psychol.*, 1956, **52**, 258–262.

CAMPBELL, B. A., & KRAELING, D. Response strength as a function of drive level and amount of drive reduction. *J. exp. Psychol.*, 1953, **45**, 97–101.

CAMPBELL, B. A., & TEGHTSOONIAN, R. Electrical and behavioral effects of different types of shock stimuli on the rat. *J. comp. physiol. Psychol.*, 1958, **51**, 185–192.

SEWARD, J. P., & PROCTER, D. M. Performance as a function of drive, reward, and habit-strength. *Amer. J. Psychol.*, in press.

SEWARD, J. P., SHEA, R. A., & DAVENPORT, R. H. Further evidence for the interaction of drive and reward. *Amer. J. Psychol.*, in press.

SEWARD, J. P., SHEA, R. A., & ELKIND, D. Evidence for the interaction of drive and reward. *Amer. J. Psychol.*, 1958, **71**, 404–407.

(Received October 5, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 4, 255-262

INFLUENCE OF EXTRATEST ILLUMINATION ON THE CRITICAL FLICKER FREQUENCY OF THE HUMAN FOVEA [1]

HARRIS RIPPS AND IRA T. KAPLAN [2]

Postgraduate Medical School, New York University

Numerous studies have shown that the critical flicker frequency (CFF) of the human fovea may be altered by illumination of neighboring regions of the retina (Berger, 1953; Fry & Bartley, 1936; Geldard, 1932; Graham & Granit, 1931; Lythgoe & Tansley, 1929). At photopic levels of a foveal test patch, the presence of a luminous area tends to produce higher values of CFF than are obtained with dark surrounds. If the inducing field [3] is contiguous with, and appreciably brighter than, the flicker area, CFF may be reduced to, or slightly below, its original value (Graham & Granit, 1931).

Several authors (Berger, 1953; Geldard, 1932; Graham & Granit, 1931) have suggested that the effects of an inducing field upon CFF are mediated by neural interaction in the retina. They assumed that the interaction occurs between the two loci of focal illumination. There is, however, the possibility that entoptic stray light, falling on and around the image of the test field, is important in determining the effects upon CFF.

A large veil of light superimposed over a smaller test field has been used previously as a substitute form of stray light (Bartley & Fry, 1934; Holladay, 1926; Stiles, 1929). It has been shown that such a veiling luminance effects changes in threshold and apparent brightness of a foveal test field similar to those produced by a peripheral inducing field (Boynton, Bush, & Enoch, 1954; Fry & Alpern, 1953).

In the present experiment veiling luminance was employed to simulate the influence of stray light on CFF. Foveal flicker thresholds were determined at various luminance levels as a function of the luminance of a veiling patch superimposed on the test field and its surround. The effect produced by the veiling luminance was analyzed into two components: the influence of the region superimposed upon the test field, and the influence of the region that surrounded the test field. In addition, results with contiguous intrafoveal test and inducing fields have been obtained over a wide range of test luminance levels.

METHOD

Apparatus.—The apparatus (Fig. 1) was a conventional dual beam photostimulator providing Maxwellian views of both target configurations. Sources S_1 and S_2 were GE 6103 concentrated filament automotive lamps connected in series and operated from a suitable battery eliminator filtered by a 12-v.

[1] This study is based upon portions of a dissertation submitted by the first author in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Faculty of Pure Science, Columbia University. The authors are indebted to C. H. Graham, C. G. Mueller, W. J. McGill, Yun Hsia, and A. Hyman for their advice and guidance.

[2] Now with System Development Corp., Lodi, New Jersey.

[3] The term *inducing field* designates any source of illumination within the stimulus configuration other than that which is presented intermittently as the test field.

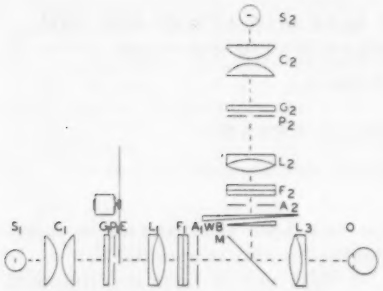

FIG. 1. Diagram of the apparatus (not to scale).

storage battery "floated" across the terminals. The monitored lamp current was maintained at slightly below rated amperage to prolong lamp life. Light from these sources was converged to form reduced images of the filaments at Ground Glasses G_1 and G_2 by Condensers C_1 and C_2. This procedure, plus the addition of apertures at P_1 and P_2, provided fairly uniform circular sources of 1 mm. diameter, which in turn were collimated by Lenses L_1 and L_2 and focused in the center of the pupillary plane of *S*'s eye (O) by L_3. The final images (d = 2 mm.) were less than the diameter of the natural pupil for conditions of maximum illuminance. The test area, illuminated by S_1, and interrupted by the rotating sector disc, E, was reduced to a hemisphere, the radius of which subtended a visual angle of 22′, by Aperture A_1. The other beam, originating at S_2 and directed to *S* by semisilvered Mirror M, was used to provide illumination of the inducing field, the form of which depended upon the aperture inserted at A_2. When the aperture was to provide an adjacent hemisphere, its dimensions were identical with those of the test area but oriented so as to have its diameter adjacent to that of the latter. The aperture for veiling luminance restricted the field to a circle subtending 3° 22′ at *S*'s eye, its center coinciding with that of the diameter of the test area. Since in certain instances these stimulus patterns were altered, the specific configuration used is indicated on each graph. Intensity control was by Wratten neutral density filters at F_1 and F_2 and a Wratten neutral density wedge and balancer, WB. Fixation was directed to the center of the test field. Head position was maintained by a dental impression bite and forehead rest. A darkroom served to shield *S* from all extraneous light.

Intermittent light stimulation was obtained by a sectored aluminum disc (E in Fig. 1) giving equal light and dark periods, and situated so as to intercept the beam at a region of narrow cross section. The flicker rate could be altered by *S* or *E* through a common friction drive of a General Radio motor speed control. This, in turn, controlled the voltage to a Bodine dc motor on the shaft of which the sector disc was mounted. Rotational velocity was read from a General Radio strobotac which was adjusted by *E* to "stop" the rotating motion of the disc following each setting made by *S*. Strobotac calibration was maintained by periodic checks against the instrument's vibrating reed during the course of each experimental session.

Stimulus luminance was determined by a monocular match of the individual Maxwellian beams to a diffuse standard, the brightness of which was measured by a Macbeth illuminometer. Luminance of the uninterrupted test source, in the absence of fixed filters was 22,800 mL.; that of the reflected beam, with the wedge and balancer set for maximum transmission, 9,800 mL.

Procedure.—Two individuals served alternately as *S* and *E* throughout the course of the experiment.

The same procedure was followed at each experimental session. A test (flicker) intensity was selected and maintained throughout. The test field, viewed monocularly, was exposed following a 10-min. period of dark adaptation. The *S* then made a preliminary series of settings for 5 to 10 min. During this period CFF decreased several cycles per second until it stabilized at a level that could be maintained for 2 hr. under constant stimulus conditions. Luminance levels of the inducing stimuli were presented in ascending order, increasing by approximately 0.5 log unit steps from near threshold visibility to the level at which either test field flicker could not be perceived or the maximum available luminance was reached.

Two responses were required of *S*: decreasing the frequency of a fused field until flicker just appears, and increasing the frequency of a flickering field until flicker just vanishes. The *E* presented fused and flickering fields in counterbalanced order.

At each luminance level of the inducing field *S* made five settings, the first of which was routinely discarded as a practice setting. Since the conditions of every experimental session were repeated, each reported measure represents the average of eight determinations. Except where otherwise indicated,

all graphs are based on the data of Subject IK. The second *S*, HR, underwent the major portion of the experiment and gave confirmatory results in all cases. Since flicker thresholds were invariably lower for this *S* (a factor which affects other characteristics of the functions), the results from the two *S*s were not averaged together.

Results

Effect of veiling luminance.—The effect of veiling luminance upon CFF at various test field luminances is shown in Fig. 2. Each curve is initially located along the CFF axis roughly in accordance with the Ferry-Porter law. At increasing levels of test field luminance, higher veiling luminances were required to alter CFF from its original value. The effect then produced was a gradual rise in CFF, increasing to maximal values as the veiling luminance was raised. As veiling luminance was further increased, CFF decreased steeply to levels far below those of the test area alone. With high veiling luminance, although flicker could be seen, the test area could not be distinguished from the veil at fusion. Encircled points are results obtained under these conditions. At still higher veiling luminances, flicker was not discernible at any frequency.

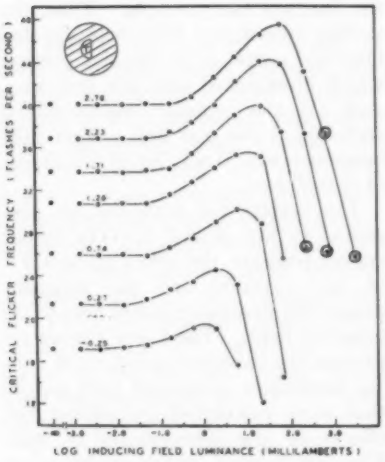

FIG. 2. CFF as a function of log inducing field luminance at various test luminance levels. The inducing field was a 3° 22′ veiling patch. Values above each curve indicate the luminance value, in log millilamberts, of the light cycle of the intermittent stimulus.

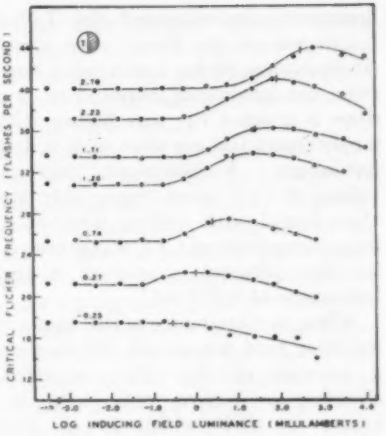

FIG. 3. CFF as a function of contiguous inducing field luminance for various test luminance levels. A short vertical line intercepts each curve at the inducing field luminance equal to the Talbot brightness of the test field alone.

Foveal CFF with steady intrafoveal stimulation.—The influence of increasing the illumination of a contiguous hemisphere is shown graphically in Fig. 3 for the same flicker luminance levels as before. Certain similarities to the effects with veiling luminance are immediately apparent. With brighter test areas, greater inducing luminance was required to raise CFF. Maximum increase of CFF was again observed at higher test luminances. As previously found by Lythgoe and Tansley (1929) and Geldard (1934), peak enhancement occurred when the inducing field

luminance approximated the Talbot brightness of the flicker area alone. At equivalent flicker luminances, however, the luminance required to produce a given CFF was higher with a contiguous inducer than with veiling luminance. Furthermore, maximal values of CFF were appreciably less than found with veiling luminance; no enhancement of CFF was produced by the adjacent source at a test luminance of 0.559 mL.

When the luminance of the adjacent inducing field was raised still further, a reversal of the effect occurred, resulting in a slow, progressive decrease in CFF which in no way resembled that produced by veiling luminance. Only with reduced intensity test fields and relatively high inducing field luminance did CFF revert to, or fall below, its initial value.

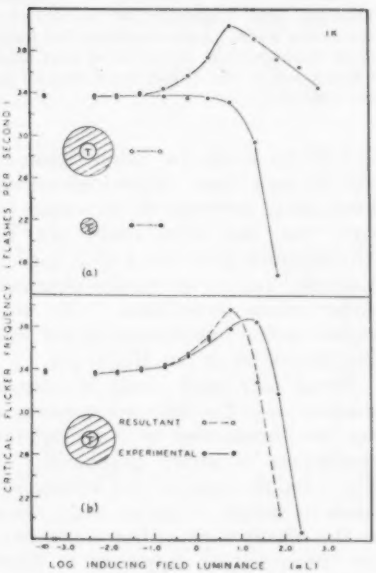

FIG. 4. CFF data of Subject IK for various inducing field configurations as a function of log luminance of the inducing field. A test field luminance of 51 mL. was used throughout. (*a*) Functions for the individual components of veiling luminance. (*b*) Results with veiling luminance compared with the results obtained by summating the effects of its component parts.

Components of veiling luminance.—The unique form of the veiling luminance function, displaying characteristics not found with contiguous stimulation, suggested an experiment in which the veiling luminance configuration was analyzed in terms of its constituent parts.

For this experiment the diameter of the veiling luminance field remained as before, 3° 22′ of visual angle, but the test area was altered to provide a circular field, 30′ in diameter. Two additional inducing fields were an area equal to that of the test field, which was superimposed upon the test field, and an annular surround equal in area to the portion of the veiling patch that exceeded the limits of the test field. Thus, the veiling luminance field was divided into two parts: (*a*) a circular portion which illuminated only the test area, and (*b*) the annular region which surrounded the test field. Test field luminance was 51 mL. for all conditions of stimulation.

The results for both *S*s are presented in Fig. 4 and 5. The upper curves indicate the effect upon CFF of the surround and the superimposed stimuli serving independently as inducing fields. The lower curves represent the experimental data of veiling luminance compared with points that were calculated by adding the effects of the components of veil. The latter were obtained by algebraically summating the changes in CFF, from its initial value, produced by each of the components at any given inducing luminance.

With a contiguous surround, both

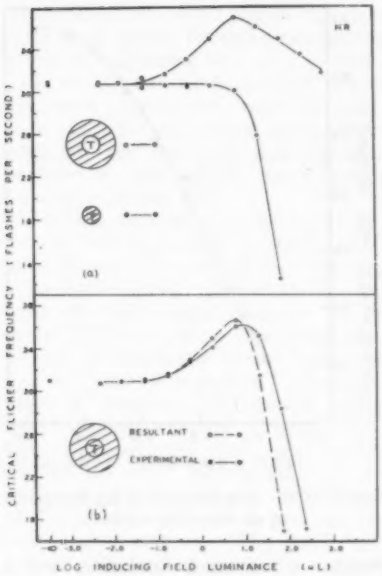

FIG. 5. CFF data of Subject HR for the same experimental conditions as Fig. 4.

the extent and rate of increase in CFF exceeded that produced by veiling luminance. The eventual decline in CFF, although proceeding more rapidly than with a small contiguous hemisphere, was still of a gradual nature, extending to surround luminance values far in excess of those obtained with veiling luminance.

When the inducing field was restricted to the test area, enhancement of CFF was completely absent; only a decrease in CFF occurred. The course of the function closely paralleled the descending limb of the curve for veiling luminance.

The curve resulting from algebraic summation of the effects produced by surround and superimposed stimuli is in close agreement with the experimentally derived veiling luminance function. However, some interaction of effects appears to exist. For example, that portion of the veiling luminance superimposed on the test field retards the enhancement effect produced by the surround component, whereas the sharp drop in CFF produced by the superimposed component is delayed by the presence of the surround. It should, however, be emphasized that these effects are small. Displacement of the experimental veiling luminance function about 0.3 log unit to the left along the abscissa leaves only small discrepancies between the two curves. Thus, synthesis of the experimental curve provides a unique demonstration of additivity of effect within the visual system.

Test field apparent brightness and CFF.—Although measures of the apparent brightness of the test field were not feasible with the present apparatus, the relation between test field apparent brightness and CFF that was observed when test field luminance was altered, was noticeably different from the relation that was found when the luminance of the inducing field was altered. In the absence of an inducing field, increasing the test field luminance resulted in a higher CFF accompanied by an increase in apparent brightness. However, when a rise in CFF was produced by a contiguous inducing field (Fig. 4a and 5a), no increase in test field apparent brightness was detectable. In fact, when the maximum enhancement of CFF occurred, the apparent brightness of the test field had decreased markedly. This decrease in brightness was so pronounced as to seem unequivocal.

DISCUSSION

The changes in CFF as a function of veiling and contiguous luminance resemble in some ways the relation of CFF to test luminance found by Hecht and

Verrijp (1933). For example, these functions exhibit an increase in CFF with increasing luminance. The curves then pass through a maximum and decline with further increase in intensity. However, several differences between these functions indicate that the inducing field effects do not result from changes within the test area similar in nature to those produced by raising the intensity of the test source. Graham and Granit (1931) have reported results with test luminance levels at the peak of the CFF vs. log I function. These authors were able to produce an increase in foveal CFF with adjacent stimulation although a further increase in test field luminance did not raise CFF. Shown graphically in Fig. 6 are the data of CFF as a function of the test field luminance for the small foveal hemisphere used previously. At high luminance levels the curve approaches a CFF value of approximately 42 flashes/sec. This value is about 5.6 cycles below the maximum obtained with veiling luminance at high test field luminance (Fig. 2). Another difference may be seen from a comparison of the slopes of these curves. Over the linear portion of the curve in Fig. 6, a log unit increase in test luminance produces an increase in CFF of 8.6 flashes/sec. The maximum increase in CFF produced by a log unit increase in inducing luminance is only 4.7 flashes/sec. (Fig. 4a). Different, too, is the luminance level required to produce a fall in CFF under the two conditions. Only when extremely high luminance levels are reached does CFF fall off with test luminance, whereas relatively low inducing luminance will cause a drop in CFF at reduced test luminance values (Fig. 2 and 3).

Analysis of the veiling luminance configuration (Fig. 4 and 5) shows that the separate effects of its component parts are responsible for the characteristic form of this function. Thus, each curve of Fig. 2 may be divided into three sections. The first, an initial horizontal portion, is found for the condition where the veiling light is of such low luminance as to exert a negligible effect upon CFF.

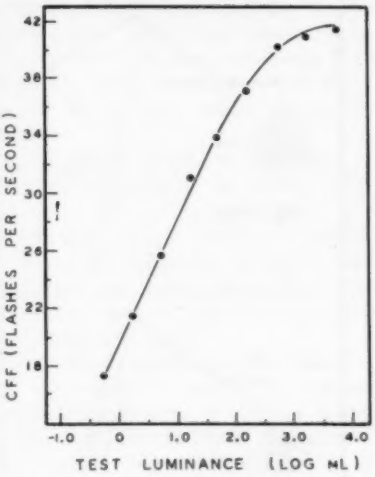

FIG. 6. CFF as a function of log luminance of the foveal test field.

Neither a surround luminance nor a steady component added to the flickering area alters CFF over this luminance range. The second section is the region in which CFF rises to a maximum and then starts to decline. Such a relationship may be considered as primarily due to the presence of a surround, since a steady component in the flickering test field does not appreciably reduce CFF when the magnitude of the steady component is no greater than that of the flickering component. The relationship in the last region, where CFF drops rapidly, reflects the effect of a steady component of large magnitude within the flickering area. This effectively masks the gradual decline in CFF produced by a relatively bright surround.

With respect to the light directly incident upon the test area, both superimposed luminance and veiling luminance may be said to constitute equivalent stimuli. The wide divergence of effects produced by these inducing fields, shown in Fig. 4 and 5, precludes the possibility that local adaptive state or the ratio of the light phase of the stimulus cycle to the level of background illumination are primary factors in the

enhancement of CFF. Since the only effect of a steady luminous component within the flickering area is to decrease CFF, the rise in CFF found with veiling luminance reflects the results of interaction between the retinal areas stimulated by the test field and by that portion of the veiling patch which surrounds the test field.

Previous experiments (Berger, 1953; Fry & Bartley, 1936; Geldard, 1932) have shown that raising the intensity of an inducing field, somewhat removed from the test locale, produces an increase in foveal CFF. This effect may be due to interaction between the two loci of focal illumination. However, physical measures of entoptic stray light (Boynton, Enoch, & Bush, 1954) have demonstrated that light falling outside the focal regions may not be neglected when considering the influence of illumination of one region of the retina upon another. The results obtained with veiling luminance suggest the hypothesis that stray light falling around the test region raises CFF. The interaction necessary for the enhancement of CFF would occur then in a region more proximal to the test locale. This local interaction may supplement direct interaction between the two areas of focal illumination, or even produce the observed effect in the absence of such direct interaction.

Summary

Foveal CFF was studied as a function of inducing field luminance for several different configurations of extratest stimulation. The inducing field effects were determined over a wide range of test field luminances with two *Ss*.

One inducing field configuration was a veiling patch superimposed upon the test field and its surrounding regions. At low veiling luminances no change in CFF was observed. As veiling luminance increased beyond a critical value, CFF gradually rose to a maximum and then fell steeply to levels far below that obtained in the absence of the veil. An inducing field adjacent to the test field also produced an increase in CFF followed by a decrease. However, the decrease in CFF was neither as steep nor as large as that produced by the veil. Even the highest luminances of the adjacent inducing field did not reduce CFF appreciably below its original value. For both inducing field configurations, at higher test field luminances the critical inducing field luminance required to raise CFF increased, and the slope and magnitude of the rise in CFF increased.

Analysis of the component parts of the veiling luminance configuration indicated that illumination of the region surrounding the test area was responsible for the increase in CFF. The only effect of a steady luminous field superimposed upon the test field was to decrease CFF. The decrease was slight until an abrupt drop occurred at high inducing luminance. Thus, the sharp decline in CFF which distinguished the effect of a veil from that of an adjacent inducer was a consequence of that portion of the veil which was superimposed upon the test field. The enhancement of CFF results therefore from interaction between adjacent retinal regions.

Other investigators have observed that CFF can be raised by an inducer somewhat removed from the test field. In the present study it was suggested that a veil of stray light from the remote inducer may contribute to the enhancement of CFF, since it was shown that a veiling patch can raise CFF. Thus, the enhancement effect may result from interaction between the test region and its immediate surround, even in the absence of interaction between remote areas of stimulation.

Subjective reports of apparent brightness indicate that a rise in CFF is not necessarily accompanied by a corresponding increase in apparent brightness. When CFF was maximally enhanced by a contiguous inducing field, test field apparent brightness was greatly reduced.

References

Bartley, S. H., & Fry, G. A. An indirect method for measuring stray light within the human eye. *J. Opt. Soc. Amer.*, 1934, **24**, 342–347.

Berger, C. Illumination of surrounding field and flicker fusion frequency with foveal images of different sizes. *Acta physiol. scand.*, 1953, **30**, 161–170.

Boynton, R. M., Bush, W. R., & Enoch, J. M. Rapid changes in foveal sensitivity resulting from direct and indirect adapting stimuli. *J. Opt. Soc. Amer.*, 1954, **44**, 56–60.

Boynton, R. M., Enoch, J. M., & Bush, W. R. Physical measures of stray light

in excised eyes. *J. Opt. Soc. Amer.*, 1954, **44**, 879–886.

FRY, G. A., & ALPERN, M. The effect of a peripheral glare source upon the apparent brightness of an object. *J. Opt. Soc. Amer.*, 1953, **43**, 189–195.

FRY, G. A., & BARTLEY, S. H. The effect of steady stimulation of one part of the retina upon the critical flicker frequency in another. *J. exp. Psychol.*, 1936, **19**, 351–356.

GELDARD, F. A. Foveal sensitivity as influenced by peripheral stimulation. *J. gen. Psychol.*, 1932, **7**, 185–189.

GELDARD, F. A. Flicker relations within the fovea. *J. Opt. Soc. Amer.*, 1934, **24**, 299–302.

GRAHAM, C. H., & GRANIT, R. Comparative studies on the peripheral and central retina. VI. Inhibition, summation and synchronization of impulses in the retina. *Amer. J. Physiol.*, 1931, **98**, 664–673.

HECHT, S., & VERRIJP, C. D. Intermittent stimulation by light. III. The relation between intensity and critical fusion frequency for different retinal locations. *J. gen. Physiol.*, 1933, **17**, 251–268.

HOLLADAY, L. L. The fundamentals of glare and visibility. *J. Opt. Soc. Amer.*, 1926, **12**, 271–319.

LYTHGOE, R. J., & TANSLEY, K. The relation of the critical frequency of flicker to the adaptation of the eye. *Proc. Roy. Soc. Lond., Ser. B*, 1929, **105**, 60–92.

STILES, W. S. The effect of glare on the brightness difference threshold. *Proc. Roy. Soc. Lond., Ser. B*, 1929, **104**, 322–351.

(Received October 8, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 4, 263–264

SUPPLEMENTARY REPORT: EFFECTIVENESS OF FOOD AND NONFOOD SIGNS IN REVERSAL LEARNING BY MONKEYS[1]

J. M. WARREN

Yerkes Laboratories of Primate Biology

Riopelle and Copelan (1954) found that rhesus monkeys learn to make errorless discrimination reversals when reversal of the stimulus-reward relationship in discrimination learning is signaled by a change in the color of the tray on which the discriminanda are presented. However, their 5 *Ss* were all trained to reverse to the tray cue. The purpose of the present experiment was to compare directly the effectiveness of failure of reinforcement and changing the color of the test tray as reversal cues for monkeys.

Method.—The 17 adolescent rhesus monkeys used as *Ss* had previously served in discrimination (Warren, 1959) and partial discrimination reversal experiments (Sinha, 1958; Warren & Sinha, 1959). They had no prior experience with problems involving direct reversals between stimuli presented simultaneously.

Apparatus, stimuli, and trial procedures were essentially the same as in Riopelle and Copelan's experiment. The stimuli were presented on a white tray to all *Ss* during the discrimination phase of the problems, and the white tray was retained during the reversal phase for the control group of 9 *Ss*. For the experimental group of 8 *Ss*, however, reversal was signaled by replacing the white tray with a black one before the first reversal trial.

The *Ss* were trained on four discrimination reversal problems each day for 20 days, a total of 80 problems. The criterion was 11 correct out of 12 noncorrection trials, with the last 8 all correct.

Results and discussion.—The significant results are displayed in Fig. 1, which shows the percentage of correct responses by the Experimental and Control Groups on postreversal Trials 1, 2, and 3 on consecutive blocks of 20 problems. Overall, the control group made less than 4% correct responses on the first trial following reversal. The experimental group made 35% and 86% correct responses on postreversal Trial 1 on the first and last problem blocks, respectively; rapid learning to reverse when the color of the test tray was changed is obvious.

A most important comparison is that between the performance of the groups on the trial following the reversal cue, Trial 1 for the experimental group vs. Trial 2 for the control group. Figure 1 shows that the control group's performance on Trial 2 surpassed the experimental group's performance on Trial 1 throughout the experiment, although the difference between the two curves diminishes on successive problem blocks. Mann-Whitney *U* tests indicate that the difference in favor of the control group is significantly greater than chance ($P < .02$) on the first three problem blocks, but that the slight advantage of the control group, less than 6%, on the last problem block is not statistically significant.

The two groups do not differ significantly in terms of success on Trials 2 or 3 following reversal. As might be inferred from the almost perfect performance of both groups on Trial 3, these sophisticated monkeys learned all of the problems to criterion very quickly. Excluding Trial 1, the monkeys averaged 0.4 errors to criterion on both the discrimination and reversal phases over the entire experiment; the means for the experi-

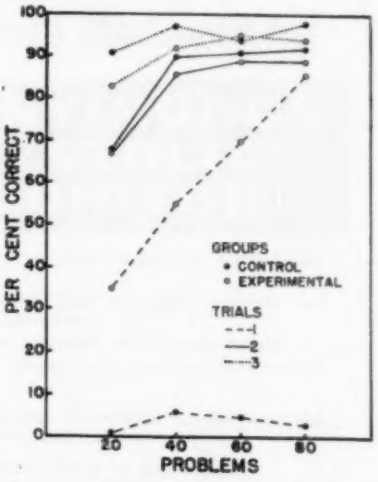

FIG. 1. Performance of the experimental and control groups on the first three trials following reversal.

[1] This research was supported by Grant M-835 from the National Institute of Mental Health, United States Public Health Service.

mental and control groups were virtually identical.

These results indicate that changing the color of the test tray becomes almost as effective, after only 60 problems, as failure of reinforcement as a reversal cue for sophisticated monkeys. The major difference between the present findings and those of Riopelle and Copelan lies in the rate with which *Ss* learned discrimination reversal to a sign. The experimental group in this study averaged 86% correct responses on the first reversal trial on Problems 61 to 80; Riopelle and Copelan's *Ss* made less than 60% correct responses at the comparable point in their training, and did not exceed 80% correct until they had been tested on 160 problems. The difference in rate of learning probably reflects differences in the previous testing experiences of *Ss*. Riopelle and Copelan's monkeys formed learning sets from the minimum possible number of stimuli (Riopelle, 1955), but *Ss* studied here had been trained on reversal-like tasks with a wider variety of stimulus materials (Sinha, 1958; Warren, 1959; Warren & Sinha, 1959).

References

RIOPELLE, A. J. Learning sets from minimum stimuli. *J. exp. Psychol.*, 1955, **49**, 28–32.

RIOPELLE, A. J., & COPELAN, E. L. Discrimination reversal to a sign. *J. exp. Psychol.*, 1954, **48**, 143–145.

SINHA, M. M. Effects of reinforcing single stimuli upon subsequent discrimination learning by rhesus monkeys. *J. comp. physiol. Psychol.*, 1958, **51**, 87–91.

WARREN, J. M. Solution of object and positional discriminations by rhesus monkeys. *J. comp. physiol. Psychol.*, 1959, **52**, 92–93.

WARREN, J. M., & SINHA, M. M. Interactions between learning sets in monkeys. *J. genet. Psychol.*, 1959, **95**, 19–25.

(Received February 3, 1960)

". . . a valuable and provocative book. It deserves to be read and discussed."

—Charles N. Cofer, Professor of Psychology, New York University

TACTICS of SCIENTIFIC RESEARCH

Evaluating Experimental Data in Psychology

by **Murray Sidman**

Here is a serious attempt to systematize and classify some of the most commonly practiced procedures in experimental sciences. Dr. Sidman, research psychologist at the Walter Reed Army Institute of Research, points out the guideposts for the evaluation of data, clarifies the techniques of verifying results, categorizes the handling of variable factors and, finally, identifies the basic characteristics of experimental design applicable not only to psychology but to all sciences.

September; $7.50

BASIC BOOKS, PUBLISHERS, 59 Fourth Avenue, New York 3, N.Y.

New Advance in Blood Pressure Measurement

CONTINUOUS SYSTOLIC MONITOR

- Accurate *continuous* measurement of blood pressure
- No arterial puncture—adult, pediatric and animal application
- Measures response to emotional and external stimuli
- Manometer pressure gauge plus connection for auxiliary recorder

Now in use in pharmacology laboratories, psychological and pediatric research, operating and recovery rooms.

Write for Brochure CSM-4 or Demonstration Representatives throughout U.S. and Canada

BIOPHYSICAL ELECTRONICS, INC.
NEW HOPE, PA.

Wiley BOOKS

Now available—

EXPERIMENTAL PSYCHOLOGY

By BURTON G. ANDREAS,
Associate Professor of Psychology,
University of Rochester.

A new text which . . .

- Promotes student growth toward research competence
- Suggests possible directions for future research
- Offers broad coverage of the traditional subject areas of the field
- Can be easily adapted to the needs and requirements of individual instructors
- Describes and refers to experiments that can be easily modified for duplication in the laboratory
- Devotes two chapters to statistical methods and techniques in experimentation
- Begins by giving a general orientation to experimentation, scientific procedure, and behavioral science
- Includes a chapter on report writing which presents a "model report"
- Discusses the relationship of theory and empirical investigation
- Gives references and additional readings at the end of every chapter

1960. 595 pages. $6.95.

Send for an examination copy.

JOHN WILEY & SONS, Inc.
440 Park Avenue South
New York 16, N.Y.